The Rovinj summer school on natural exponential families

Gérard Letac*

Table des matières

*Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 31062, Toulouse, France.

A summer academy for undergraduate or young graduate German students was organized in Rovinj, Croatia in September 2004. One group was working on natural exponential families. A characteristic of the academy is the fact that lectures are only given by the students. Each student has to give a lecture 2 or 3 hours long. The instructors (Lutz Mattner, Gérard Letac) provided list of topics, and books or papers to read before the lecture. The following rhapsodic document provides help and comments.

I The variance function e^m.

G AS A LIMIT. The variance function $V_G(m) = e^m$ occurs mainly as the limit of the power variance function $V_{F_n}(m) = (1 + \frac{m}{n})^n$ for $n \to \infty$. According to the two basic rules

– JORGENSEN RULE. If $\lambda \in \Lambda(F)$ then $V_{F_\lambda}(m) = mV_F(\frac{m}{\lambda})$

– AFFINE RULE. If $\varphi(x) = ax + b$ then $V_{\varphi(F)}(m) = a^2 V_F(\frac{m-b}{a})$,

the above F_n is obtained by starting from the classical NEF on $(0, \infty)$ with variance function m^n, by applying the Jorgensen rule to $\lambda = n$ and then the affine rule to

$$\varphi(x) = n^{\frac{1}{n-2}} x - n.$$

In particular $M_{F_n} = (-n, \infty)$. Thus G exists and $M_G = \mathbb{R}$. In particular G is steep.

LAPLACE TRANSFORMS. We compute a particular $\mu \in \mathcal{B}(G)$. We have $\psi'_\mu(m) = e^{-m}$ and $\Theta(\mu) = (-\infty, 0)$. We choose $\theta = \psi_\mu(m) = -e^{-m}$, $m = k'_\mu(\theta) = -\log(-\theta)$, we choose $k_\mu(\theta) = -\theta \log(-\theta) + \theta$ and we get

$$L_\mu(\theta) = e^\theta (-\theta)^{-\theta}.$$

We observe that $D(\mu) = (-\infty, 0]$ and that $L_\mu(0) = 1$, that is to say that μ is a probability. It does not belong to $G = F(\mu)$ but it belongs to the so called "full exponential family" $\overline{G}$ generated by μ. The probability μ in particular has no exponential moments, in the sense that for all $\alpha > 0$ we have

$$\int_{\mathbb{R}} e^{\alpha|x|} \mu(dx) = \infty.$$

The part e^θ in the Laplace transform of μ may seem unpleasant. Well, replacing μ by $\mu_1 = \mu * \delta_{-1}$ ie shifting G on the left of one unit provides the more pleasant Laplace transform $L_{\mu_1}(\theta) = (-\theta)^{-\theta}$ and the less pleasant variance function $V_{G_1}(m) = e^{m-1}$. Anyway, the explicit calculation of μ_1 does not seem possible.

THE FOURIER TRANSFORMS OF THE ELEMENTS OF G_1. Let us fix $\theta_0 \in \Theta(\mu_1) = (-\infty, 0)$. The Laplace transform of $P(\theta_0, \mu_1)$ is

$$\frac{(-s - \theta_0)^{-s-\theta_0}}{(-\theta_0)^{-\theta_0}} = e^{-s\log(-\theta_0)} \frac{1}{(1 + \frac{s}{\theta_0})^{s+\theta_0}}.$$

As a function of s this function is analytic is the half complex plane

$$\{s; \Re s < -\theta_0\}.$$

thus, doing $s = it$ with $t \in \mathbb{R}$ gives the characteristic function of $P(\theta_0, \mu_1)$ as

$$e^{-it\log(-\theta_0)}\frac{1}{(1+\frac{it}{\theta_0})^{it+\theta_0}}.$$

THE FOURIER TRANSFORMS OF μ AND μ_1. Moving from the Laplace transform to the Fourier transform for the probabilities μ and μ_1 is less easy, since their Laplace transforms are analytic in the open half complex plane $\{s; \Re s < 0\}$ which does not include the imaginary axis used for the Fourier transform. Denoting $-\theta_0 = \epsilon > 0$ for simplicity, we now compute for real t $\lim_{\epsilon\to 0}(\epsilon - it)^{\epsilon - it}$. For this we assume first $t > 0$ and we observe that $-i = e^{-i\frac{\pi}{2}}$. We now introduce the analytic function $z \mapsto u = \log z$ on the complex domain $D = \mathbb{C} \setminus \{z; \Im z = 0, \Re z \leq 0\}$ such that $z = e^u$. The most important property is that $\log(zz') = \log z + \log z'$ for z, z', zz' in D. With this notation we write

$$\log(\epsilon - it) = \log t - i\frac{\pi}{2} + \log(1 + \frac{e^{i\frac{\pi}{2}}\epsilon}{t}) \to_{\epsilon\to 0} \log t - i\frac{\pi}{2}.$$

Thus

$$(\epsilon - it)^{\epsilon - it} = e^{(\epsilon - it)\log(\epsilon - it)} \to_{\epsilon\to 0} e^{-it(\log t - i\frac{\pi}{2})} = e^{-t\frac{\pi}{2}}e^{-it\log t)}.$$

Using the Hermitian symmetry of characteristic functions, the Fourier transform of μ_1 is for all $t \neq 0$

$$e^{-|t|\frac{\pi}{2}}e^{-it\log|t|}$$

(See Feller vol 2 page 542, first edition). The Fourier transform of μ is $e^{-|t|\frac{\pi}{2}}e^{it(1-\log|t|)}$. An immediate consequence is that for X and Y independent with the same distribution μ_1 then the distribution of $Z = X - Y$ is the Cauchy distribution $\frac{2dz}{4z^2+\pi^2}$ since the characteristic function of Z is $e^{-|t|\frac{\pi}{2}}$. An other important feature of the characteristic function $e^{-|t|\frac{\pi}{2}}e^{it(1-\log|t|)}$ is that it is integrable. It implies that the Fourier inversion is available and that $\mu_1(dx)$ has a density f given by the formula

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{-|t|\frac{\pi}{2}}e^{it(1-\log|t|)}dt = \frac{1}{\pi}\int_0^{\infty} e^{-t\frac{\pi}{2}}\cos(t(x - 1 + \log t))dt.$$

I do not think that explicit calculation of this density is feasible.

ELEMENTS OF $\overline{G_1}$ ARE INFINITELY DIVISIBLE. Since $k''_{\mu_1}(\theta) = \frac{1}{-\theta}$ is the Laplace transform of the positive measure $\mathbf{1}_{(0,\infty)}(x)dx$ this implies that the elements of $\overline{G_1}$ are infinitely divisible. Furthermore, this implies that the Lévy measure of $P(\theta_0, \mu_1)$ is $\nu(dx) = e^{\theta_0 x}\frac{1}{x^2}\mathbf{1}_{(0,\infty)}(x)dx$. One notices that $\nu(dx)$ is of type 2, that is $\int_{\mathbb{R}\setminus\{0\}}\min(1,\, x^2)\nu(dx) < \infty$ as any Lévy measure, but $\nu(dx)$ is such that

$$\int_{\mathbb{R}\setminus\{0\}}\min(1,\, |x|)\nu(dx) = \infty.$$

This implies in particular that although the Lévy measure of $P(\theta_0, \mu_1)$ is concentrated on $(0, \infty)$, however $P(\theta_0, \mu_1)$ is not concentrated on $(0, \infty)$ but has $\mathbb{R}$ as support interval. The Laplace transform of $P(\theta_0, \mu_1)$ can also be written

$$e^{-s\log(-\theta_0)}\frac{1}{(1+\frac{s}{\theta_0})^{s+\theta_0}} = \exp(a_\tau s + \int_0^\infty (e^{sx} - 1 - s\tau(x))e^{\theta_0 x}\frac{dx}{x^2})$$

where τ is any bounded function on $\mathbb{R} \setminus \{0\}$ such that $\tau(x) - x = o(x)$ around 0, and a_τ depends on the chosen τ. This shows that any choice of τ is rather artificial. For an arbitrary Lévy measure ν, the move to τ to an other τ_1 is given by

$$a_\tau - a_{\tau_1} = \int_{\mathbb{R}\setminus\{0\}} (\tau_1(x) - \tau(x))\nu(dx).$$

For instance Feller vol 2 uses $\tau(x) = \sin x$ and the Russian literature uses $\tau(x) = x/(1+x^2)$.

For the Lévy measure $\nu(dx) = e^{\theta_0 x}\frac{1}{x^2}\mathbf{1}_{(0,\infty)}(x)dx$, the computation of a_τ for these classical τ does not seem possible (Feller on page 543 computes something different). But choosing $\tau(x) = xe^{-|x|}$ leads to $a_\tau = -\log(1 - \theta_0)$. For this we use the Frullani integral

$$\int_0^\infty (f(ax) - f(bx))\frac{dx}{x} = f(0)\log\frac{b}{a}$$

where f is continuous on $[0, \infty)$ such that $\int_1^\infty |f(x)|\frac{dx}{x} < \infty$ and where $a, b > 0$. Applying it to $f(x) = e^{-x}$ and to $a = 1 - \theta_0$ and $b = -s - \theta_0$ leads the result.

$-G$ HAS A RECIPROCAL FAMILY. Consider the image $-G$ of G by the map $x \mapsto -x$. Its variance function is e^{-m}. Its Lévy measure is concentrated on the negative line, thus $-G$ admits a reciprocal NEF (say A) whose domain of the means is $M_A = (0, \infty)$ and whose variance function is

$$V_A(m) = m^3 e^{-\frac{1}{m}}.$$

Thus A is the family of distributions of the stopping time $T = \inf\{t; X(t) = 1\}$ where X is a Lévy process such that the distribution of $-X(1)$ is $P(\theta_0, \mu)$ when we let θ_0 vary in $(-\ infty, 0)$. See Letac Mora 1992 for details about reciprocity. A generating measure α of A cannot be computed explicitely. Replacing $-G$ by a translate G_b with variance function e^{-m+b} replaces A by A_b with variance function $M_{A_b} = (0, \infty)$

$$V_{A_b}(m) = m^3 e^{-\frac{1}{m}+b} = e^b V_A(m)$$

on $M_{A_b} = (0, \infty)$. This makes a marked difference among other NEF and could be formulated in a sort of (rather trivial) characterization of G.

II The variance function $t(1 + \frac{m^2}{t^2})$

Theorem. Let $t > 0$. The natural exponential family F_t with domain of the means $\mathbb{R}$ and variance function $t(1 + \frac{m^2}{t^2})$ is generated by the probability

$$\mu_t(dx) = \frac{2^{t-2}}{\pi}\left|\Gamma(\frac{t + ix}{2}\right|^2 \frac{1}{\Gamma(t)}dx.$$

Its Laplace transform is defined on $\Theta(\mu_t) = (-\frac{\pi}{2}, \frac{\pi}{2})$ and is $(\cos\theta)^{-t}$. In particular, for $t = 1, 2$ we have

$$\mu_1(dx) = \frac{1}{2\cosh\frac{\pi x}{2}} \quad \mu_2(dx) = \frac{x}{2\sinh\frac{\pi x}{2}}.$$

Proof. We adopt the notations of the Letac-Mora paper (1991) : if μ is a positive non Dirac Radon measure on $\mathbb{R}$ consider its Laplace transform

$$L_\mu(\theta) = \int_{-\infty}^{\infty} e^{\theta x}\mu(dx) \leq \infty.$$

Assume that the interior $\Theta(\mu)$ of the interval $D(\mu) = \{\theta \in \mathbb{R}; L_\mu(\theta) < \infty\}$ is not empty. Write $k_\mu = \log L_\mu$. Then the family of probabilities

$$F = F(\mu) = \{P(\theta, \mu)\ ;\ \theta \in \Theta(\mu)\}$$

where

$$P(\theta, \mu)(dx) = e^{\theta x - k_\mu(\theta)}\mu(dx)$$

is called the natural exponential family (NEF) generated by μ. Note that $F(\mu) = F(\nu)$ if and only if there exists a and b such that $\nu(dx) = e^{ax+b}\mu(dx)$. This implies that $F(\mu)$ can be generated by one of its members as well as sometimes an unbounded measure.

Two basic results are $k'_\mu(\theta) = \int_{-\infty}^{\infty} xP(\theta, \mu)(dx)$ and the fact that k'_μ is increasing (or that k_μ is convex. The set $k'_\mu(\Theta(\mu)) = M_F$ is called the domain of the means. We denote by $\psi_\mu : M_F \to \Theta(\mu)$ the reciprocal function of k'_μ. Thus $F(\mu)$ can be parametrized by M_F by the map from M_F to F which is

$$m \mapsto P(\psi_\mu(m), \mu) = P(m, F).$$

One can prove that the variance $V_F(m)$ of $P(m, F$ is

$$V_F(m) = k''_\mu(\psi_\mu(m)) = \frac{1}{\psi'_\mu(m)}. \tag{2.1}$$

The map $m \mapsto V_F(m)$ from M_F to $(0, \infty)$ is called the variance function and characterizes F.

Let us apply these concepts to the finding of a generating measure μ for the variance function $t(1 + \frac{m^2}{t^2})$. We use 2.1 for writing with $m = tu$

$$d\theta = \psi'_\mu(m)dm = \frac{dm}{t(1 + \frac{m^2}{t^2})} = \frac{du}{1 + u^2} = d\arctan u.$$

Thus $m = t\tan\theta = k'_\mu(\theta)$ and $k_\mu(\theta) = -t\log\cos\theta$ on $\Theta(\mu) = (-\frac{\pi}{2}, \frac{\pi}{2})$. The remark about $F(\mu) = F(\nu)$ leads us to ignore the two integration constants in the process. Note that here $k_\mu(0) = 0$ implies that the chosen μ is a probability if it exists.

The interesting point is now to prove that μ indeed exists and to compute it. For this the analytic character of the Laplace transform enables us to declare that the Fourier transform of μ is

$$s \mapsto \varphi(s) = \frac{1}{(\cosh s)^t}.$$

Luckily, this is an integrable function and the Fourier inversion formula applies : the density of μ should exists and should be

$$f_t(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \frac{1}{(\cosh s)^t} ds.$$

The trick to compute this integral in a elementary way is to write $v = e^{2s}$. We get

$$f_t(x) = \frac{2^{t-2}}{\pi} \int_0^{\infty} \frac{v^{\frac{t-ix}{2}-1} dv}{(1+v)^t}$$

We rely now on two formulas about the Gamma function

$$\int_0^{\infty} \frac{v^{a-1} dv}{(1+v)^{a+b}} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad \Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin \pi p},$$

where the real or complex numbers $a, b, p, 1-p$ have a positive real part. Applying the first formula to $a = \frac{t-ix}{2}$ and to $b = \bar{a} = \frac{t+ix}{2}$ and using the fact that $\Gamma(a) = \overline{\Gamma(\bar{a})}$ we get the announced value of $f_t(x)dx = \mu_t(dx)$ which is indeed positive. To get μ_1 use $\Gamma(p)\Gamma(1-p) = \frac{\pi}{\sin \pi p}$ with $p = \frac{1-ix}{2}$ and the sinus of a complex argument. To get μ_2 use the fact that its density is given by the convolution

$$\frac{1}{4} \int_{-\infty}^{\infty} \frac{dy}{\cosh \frac{\pi(x-y)}{2} \cosh \frac{\pi y}{2}}.$$

This integral is easily computed by the change of variable $v = e^{\pi y}$ since we are led to the integral of a rational fraction of v.

III The Poincaré characterization of natural exponential families

Theorem. Let $I = (a, b)$ be an open interval of the real line. Let $\nu(dx)$ be a positive measure on I. We assume that $\nu(a, a+\epsilon)$ and $\nu(b-\epsilon, b)$ are positive for all $\epsilon > 0$. Let $(x, m) \mapsto f(x, m)$ be a real function of class C^2 on I^2 such that $P_m(dx) = e^{f(x,m)}\nu(dx)$ is a probability on I for all m. Assume that $m \mapsto P_m$ is not constant. If $(x_1, \ldots, x_n) \in I^n$ we write $\overline{x}_n = \frac{1}{n}(x_1 + \cdots + x_n)$.

Let us assume that for all $n \geq 1$ if $X_1, \ldots, X_n$ are iid with distribution P_m with unknown m then $\overline{X}_n$ is a maximal likelihood estimator of m. Then there exist three real C^2 functions a, b, c such that $f(x, m) = xa(m) + b(m) + c(x)$. Finally $F = \{P_m\,; m \in I\}$

is the natural exponential family generated by $e^{c(x)}\nu(dx)$. It satisfies $\int_I xP_m(dx) = m$ for all $m \in I$. It is a steep family with I as domain of the means.

Proof. Saying that $\overline{X}_n$ is the maximal likelihood estimator of m is saying that for all $(x_1, \dots, x_n) \in I^n$ the maximum of the function $m \mapsto \sum_{i=1}^n f(x_i, m)$ is reached on $m = \overline{x}_n$ which implies that the derivative is zero on this point. Thus we write for all $(x_1, \dots, x_n) \in I^n$

$$\sum_{i=1}^n \frac{\partial}{\partial m} f(x_i, \overline{x}_n) = 0.$$

Now we derive the above expression with respect to x_j. For simplicity we write $g(x, m) = \frac{\partial^2}{\partial x \partial m} f(x, m)$. Since $\frac{\partial}{\partial x_j} \overline{x}_n = \frac{1}{n}$ we get

$$g(x_j, \overline{x}_n) + \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial m^2} f(x_i, \overline{x}_n) = 0.$$

Thus $j \mapsto g(x_j, \overline{x}_n)$ does not depend on j. For instance we get that for all $(x_1, \dots, x_n) \in I^n$ we have

$$g(x_1, \overline{x}_n) = g(x_2, \overline{x}_n). \tag{3.2}$$

Since 3.2 holds for all n, given (x_1, x_2, m) in I^3 one can always find $n \geq 3$ and $(x_3, \dots, x_n)$ in I^{n-2} such that $m = \overline{x}_n$ holds. For this, consider the number $y_n = m + \frac{2m - x_1 - x_2}{n-2}$. Since I is open, there exists n big enough such that $y_n \in I$ and we shall take $x_3 = \dots = x_n = y_n$. Thus there exists a continuous function $a'(m)$ such that $g(x, m) = a'(m)$ for all x from which we get $\frac{\partial}{\partial x} f(x, m) = a(m) + c'(x)$ for some continuous function c' and $f(x, m) = xa(m) + b(m) + c(x)$ for some function b.

Consider now the measure $\mu(dx) = e^{c(x)}\nu(dx)$. Observe that $m \mapsto a(m)$ is not a constant function. If we had $a(m) = a$ for all m we would get $\int_I e^{ax}\mu(dx) = e^{-b(m)}$ for all $m \in I$, $m \mapsto b(m)$ would be a constant and P_m would not depend on m, a contradiction. Consider the interval $J = a(I)$. We just have seen that it has a non empty interior $\dot{J}$ and this implies that $\Theta(\mu)$ is not empty. Thus for all $m \in I$ we have $k_\mu(a(m)) = -b(m)$. Now, since the hypothesis holds for $n = 1$ we have $f(x, m) \leq f(x, x)$ which implies $\frac{\partial}{\partial m} f(x, x) = xa'(x) + b'(x) = 0$. Thus taking derivative of $k_\mu(a(x)) = -b(x)$ we get $k'_\mu(a(x)) = x$ for all $x \in I$. This implies that a is injective and that the image J of I is open. Thus a is the restriction of ψ_μ to I. Now, by definition, the domain of the means of $F(\mu)$ is contained in I. Thus they coincide and $\int_I xP_m(dx) = m$ for all $m \in I$. The coincidence of the domain of the means and of the interior of of the closed convex support is one of the definitions of steepness. Thus $F = F(\mu)$ is steep.

IV An unpublished paper with V. Seshadri

Infinite divisibility of the hitting time of a right continuous random walk :

Theorem. Let $(X_n)_{n \geq 1}$ be a sequence of non Dirac iid random variables valued in

$\{1, 0, -1, \ldots\}$ such that $0 \le \mathbb{E}(X_1) \le \infty$. Then $T = \inf\{n;\ X_1 + \cdots + X_n = 1\}$ is infinitely divisible.

Proof. The fact that $0 \le \mathbb{E}(X_1)$ implies $p_1 > 0$ and implies that $\Pr(T < \infty) = 1$ from classical properties of random walks (see Spitzer (1964)). Denote $p_n = \Pr(X_1 = n)$, $S_n = X_1 + \cdots + X_n$ and for $|z| \le 1$ $g(z) = \sum_{n=0}^{\infty} p_{-n+1} z^n = \mathbb{E}(z^{1-X_1})$. Consider the martingale $n \mapsto M_n(z) = z^{n-S_n} g(z)^{-n}$ with respect to the natural filtration associated to $(X_n)_{n\ge 1}$. Since T is a regular stopping time with respect to this filtration it is well known that $\mathbb{E}(M_T(z)) = 1$ (see for instance Letac and Mora (1990), Th. 5.6). This is rewritten as $\mathbb{E}(w^T) = z$ where $w = z/g(z)$. The Lagrange formula, under the form stated in Letac and Mora (1990), Th. 4.1 implies the existence of a positive R and the existence of an analytic function $w \mapsto h(w)$ on the disc $D_R = \{w; |w| < R\}$ valued in the unit disc D_1 such that $h(w) = wg(h(w))$. Furthermore, for any function F analytic in D_1 we have

$$F(h(w)) = F(0) + \sum_{n=1}^{\infty} \frac{w^n}{n!} \left[\left(\frac{d}{dz} \right)^{n-1} (F'(z)(g(z))^n \right]_{z=0}. \tag{4.3}$$

Thus $\mathbb{E}(w^T) = h(w)$ is the generating function of the random variable T. Note that T is concentrated on $\{1, 2, \ldots\}$ and that trivially $\Pr(T = 1) = p_1$. We denote $\alpha = -\log p_1$. Proving the infinite divisibility of T is equivalent to proving the infinite divisibility of $T-1$, whose generating function is $h(w)/w$. From Feller (1968) page 290 this is equivalent to prove that the analytic function f defined in D_R by

$$f(w) = 1 + \frac{1}{\alpha} \log \frac{h(w)}{w} = \frac{1}{\alpha} \log \frac{h(w)}{p_1 w} = \frac{1}{\alpha} \log \frac{1}{p_1 g(h(w))}$$

has a power series expansion with non negative coefficients. Let us now consider $F(z) = 1 - \frac{p_1}{g(z)}$. It satisfies $F(0) = 0$ and

$$f(w) = \frac{1}{\alpha} \log \frac{1}{1 - F(h(w))} = \frac{1}{\alpha} \sum_{k=1}^{\infty} \frac{1}{k} (F(h(w))^k.$$

From this formula, in order to prove that f has non negative coefficients, it is enough to prove that $w \mapsto F(g(w))$ has non negative coefficients. This is easily achieved by 4.3 : here for $n \ge 1$

$$\left[\left(\frac{d}{dz} \right)^{n-1} (F'(z)(g(z))^n) \right]_{z=0} = p_1 \left[\left(\frac{d}{dz} \right)^{n-1} (g'(z)(g(z))^{n-2}) \right]_{z=0},$$

which is clearly non negative, since g has non negative coefficients.

References :

Feller, W. (1968) *An Introduction to Probability Theory and Its Applications, Vol 1, 3rd Ed.*, Wiley, New York .

Letac, G., Mora, M. (1990) " Natural Exponential Families with Cubic Variance Functions", *Ann. Statist.* **18**, 1-37.

Spitzer, F. (1964) *Principles of Random Walk*, Van Nostrand, Princeton.

V Using the above result for simplifying a proof.

In the Letac-Mora paper of 1990, the proof of the following result is rather horrible

Proposition. Let $t > 0$. Given a NEF F_1 concentrated on nonnegative integers such that 0 is charged by F_1 then there exists a NEF G_t concentrated on nonnegative integers such that 0 is charged by G_t and such that the variance functions are related by

$$V_{G_t}(m) = \frac{(m+t)^3}{t^2} V_{F_1}(\frac{m}{m+t}).$$

in a neighborhood of 0.

Proof. We prove it first for $t = 1$. Denote by $(0, b_1)$ the domain of the means of F_1. Let $(\nu_n)_{n\geq 0}$ be a member of F_1. Since F_1 charges 0 there exists such a μ such that its expectation is < 1. Now consider the image F_2 of F_1 by $x \mapsto 1 - x$. Then the variance function of F_2 is $V_{F_2}(m) = V_{F_1}(1 - m)$ with domain of the means $(1 - b_1, 1)$. Now the reciprocal family F_3 of F_2 exists and is concentrated on positive integers. Its domain of the means is $(1, \infty)$ if $b_1 \geq 1$ and is $(1, \frac{1}{1-b_1}$ if $b_1 < 1$. Its variance function is $V_{F_3}(m) = m^3 V_{F_2}(1/m) = m^3 V_{F_1}((m-1)/m)$. Now define G_1 as the image of F_3 by $x \mapsto x - 1$. Its domain of the means is $(0, \infty)$ or $(0, \frac{b_1}{1-b_1}$ according to $b_1 \geq 1$ or not. Its variance function is

$$V_{G_1}(m) = V_{F_3}(m+1) = (m+1)^3 V_{F_1}(\frac{m}{m+1}).$$

Now to pass to a general $t > 0$ we use the preceeding result which says that actually F_3 is infinitely divisible. This G_1 is also infinitely divisible, and there exists a NEF G_t with variance function

$$V_{G_t}(m) = tV_{G_1}(\frac{m}{t}) = t(\frac{m}{t}+1)^3 V_{F_1}(\frac{\frac{m}{t}}{\frac{m}{t}+1}) = \frac{(m+t)^3}{t^2} V_{F_1}(\frac{m}{m+t}).$$

VI The Wishart integral

(Apologies : this was written in French)

Nous calculons pour commencer une intégrale auxiliaire :

Proposition 1. Soit E et F deux espaces euclidiens de dimensions n et m, et soit a et b des endorphismes symétriques définis positifs de E et F respectivement. On munit $L = L(E, F)$ de la structure euclidienne $\langle x, y\rangle = \mathrm{tr}\,(x^*y)$, où $x^* \in L(F, E)$ est l'adjoint de x, et on munit L de la mesure de Lebesgue dx associée à cette structure euclidienne. Alors

1. L'application $x \mapsto bxa$ est un endomorphisme symétrique défini positif de L, de déterminant $D = \det a^m \det b^n$.
2. Pour tout $y \in L$ on a

$$\int_L e^{-\mathrm{tr}\,(x^* bxa)+2\,\mathrm{tr}\,(x^* y)} dx = \frac{\pi^{mn/2}}{D^{1/2}} e^{\mathrm{tr}\,(y^* b^{-1} y a^{-1})}. \tag{6.4}$$

3. La fonction sur L définie par $x \mapsto e^{-\mathrm{tr}\,(x^* bxa)+2\,\mathrm{tr}\,(x^* y)}$ est proportionnelle à la densité d'une loi gaussienne sur L de covariance $x \mapsto \frac{1}{2} b^{-1} x a^{-1}$ et de moyenne $b^{-1} y a^{-1}$.

Démonstration. Puisque $(y^* bxa)^* = ax^* by$ il est clair que $\langle y, bxa \rangle = \langle x, bya \rangle$ et donc que $x \mapsto bxa$ est un endomorphisme symétrique de L. Soit e et f des bases orthonormées de E et F qui diagonalisent a et b. Notons $[a]_e^e = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ et $[b]_f^f = \mathrm{diag}(\beta_1, \ldots, \beta_m)$.

Si $x \in L$ a pour matrice représentative $[x]_e^f = (x_{ij})$ alors pour $x \neq 0$ on a

$$\mathrm{tr}\,(x^* bxa) = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_j \beta_i x_{ij}^2 > 0, \tag{6.5}$$

ce qui montre la positivité de $x \mapsto bxa$. Il est clair que ses vecteurs propres sont les $f_i \otimes e_j$, lélément de L défini par $v \mapsto f_i \langle e_j, v \rangle$, et qui est associé à la valeur propre $\alpha_j \beta_i$. Donc son déterminant est

$$D = \prod_{ij} (\alpha_j \beta_i) = \left(\prod_{i=1}^{m} \beta_i \right)^n \left(\prod_{j=1}^{n} \alpha_j \right)^m = (\det a)^m (\det b)^n.$$

Ensuite, pour exploiter 6.5) on écrit

$$\int_{-\infty}^{\infty} e^{-\alpha_j \beta_i x_{ij}^2 + 2 x_{ij} y_{ij}} dx_{ij} = \frac{\pi^{1/2}}{(\alpha_j \beta_i)^{1/2}} e^{\frac{y_{ij}^2}{\alpha_j \beta_i}}.$$

En utilisant $\sum_{i=1}^m \sum_{j=1}^n \frac{y_{ij}^2}{\alpha_j \beta_i} = \mathrm{tr}\,(y^* b^{-1} y a^{-1})$ ainsi que 6.5) on a le résultat 6.4). La troisième partie est une conséquence immédiate de 6.4).

Proposition 2. Si P_r est le cône des matrices symétriques d'ordre r définies positives et si $p > (r-1)/2$ alors pour $\theta \in P_r$ on a

$$\int_{P_r} \exp -\mathrm{tr}\,(\theta x^{-1}) (\det x)^{-p-\frac{r+1}{2}} dx = \int_{P_r} \exp -\mathrm{tr}\,(\theta y) (\det y)^{p-\frac{r+1}{2}} dy = \frac{\Gamma_r(p)}{(\det \theta)^p}. \tag{6.6}$$

où dy est la mesure de Lebesgue[1] sur P_r définie par $dy = \prod_{1 \leq i \leq j \leq r} dy_{ij}$ et où

$$\Gamma_r(p) = \pi^{\frac{r(r-1)}{4}} \prod_{j=1}^{r} \Gamma(p - \frac{j-1}{2}). \tag{6.7}$$

[1]C'est l'usage des statisticiens que nous conservons ici plutôt que de prendre la mesure de Lebesgue canonique pour la structure euclidienne engendrée par $\mathrm{tr}\,(yy^T)$.

Remarque. Si $p > (r-1)/2$, $\theta \in P_r$ et $\sigma \in P_r$ les probabilités sur P_r

$$\frac{(\det\theta)^p}{\Gamma_r(p)}\exp -\mathrm{tr}\,(\theta x^{-1})(\det x)^{-p-\frac{r+1}{2}}\mathbf{1}_{P_r}(x)dx, \tag{6.8}$$
$$\frac{1}{(\det\sigma)^p\Gamma_r(p)}\exp -\mathrm{tr}\,(\sigma^{-1}y)(\det y)^{p-\frac{r+1}{2}}\mathbf{1}_{P_r}(y)dy$$

sont respectivement appelées loi inverse Wishart ordinaire de paramètres p et θ, et loi de Wishart ordinaire de paramètres p et σ. Il est évident que si $X = Y^{-1}$ alors Y suit une loi de Wishart ordinaire de paramètres (p,σ) si et seulement si X suit une loi inverse Wishart ordinaire de paramètres (p,θ) avec $\theta = \sigma^{-1}$.

Démonstration. La première égalité de 6.6) découle du changement de variable $y = x^{-1}$ dont la différentielle est $h \mapsto -x^{-1}hx^{-1}$, un endomorphisme de l'espace des matrices réelles symétriques d'ordre r. En imitant la démonstration de la partie 1 de la Proposition 1, on voit que le déterminant de cet endomorphisme est $(-1)^r(\det x)^{-r-1}$, ce qui permet de calculer le jacobien de $x \mapsto x^{-1}$ et de montrer cette première égalité. Puisque la démonstration de Bartlett de la deuxième égalité de 6.6) figure dans tous les ouvrages d'analyse multivariée, nous allons en donner une autre. On procède par récurrence sur r. Pour $r = 1$, c'est la définition de la fonction gamma. Supposons le résultat vrai pour tous les entiers $< r$ et soit m et n des entiers > 0 tels que $m + n = r$. Ecrivons alors y et θ dans P_r ainsi :

$$y = \begin{bmatrix} y_1 & y_{12} \\ y_{21} & y_2 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 & \theta_{12} \\ \theta_{21} & \theta_2 \end{bmatrix}$$

avec $y_1 \in P_m$ et $y_2 \in P_n$, et notons $y_1' = y_1 - y_{12}y_2^{-1}y_{21}$. L'égalité habituelle

$$y = \begin{bmatrix} 1 & y_{12}y_2^{-1} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y_1' & 0 \\ 0 & y_2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ y_2^{-1}y_{21} & 1 \end{bmatrix},$$

montre que $y \in P_r$ si et seulement si $y_1' \in P_m$ et $y_2 \in P_n$. On a alors

$$\begin{aligned}
&\int_{P_r}\exp -\mathrm{tr}\,(\theta y)(\det y)^{p-\frac{r+1}{2}}dy\\
&\overset{(1)}{=} \int_{P_m\times P_n\times R^{nm}} e^{-\mathrm{tr}\,(\theta_1 y_1')-\mathrm{tr}\,(\theta_2 y_2)+\mathrm{tr}\,(\theta_1 y_{12}y_2^{-1}y_{21})-2\mathrm{tr}\,(\theta_{21}y_{12})}\\
&\qquad\times(\det y_1')^{p-\frac{1}{2}(r+1)}(\det y_2)^{p-\frac{1}{2}(r+1)}dy_1'dy_{12}dy_2\\
&\overset{(2)}{=} \pi^{mn/2}(\det\theta_1)^{-n/2}\int_{P_m\times P_n} e^{-\mathrm{tr}\,(\theta_1 y_1')-\mathrm{tr}\,(y_2(\theta_2-\theta_{21}\theta_1^{-1}\theta_{12}))}\\
&\qquad\times(\det y_1')^{p-\frac{1}{2}(r+1)}(\det y_2)^{p-\frac{1}{2}(n+1)}dy_1'dy_{12}dy_2\\
&\overset{(3)}{=} \pi^{mn/2}(\det\theta_1)^{-n/2}\frac{\Gamma_m(p-\frac{n}{2})}{(\det\theta_1)^{p-\frac{n}{2}}}\frac{\Gamma_n(p)}{(\det(\theta_2-\theta_{21}\theta_1^{-1}\theta_{12}))^p}\\
&\overset{(4)}{=} \pi^{mn/2}\frac{\Gamma_m(p-\frac{n}{2})\Gamma_n(p)}{(\det(\theta)^p} \overset{(5)}{=} \frac{\Gamma_r(p)}{(\det\theta)^p}.
\end{aligned}$$

Dans cette suite d'égalités, (1) vient du changement de variable $y \mapsto (y_1', y_2, y_{12})$ de jacobien 1, l'égalité (2) vient de l'intégrale par rapport à dy_{12} calculée en appliquant la Proposition 1 à $(a, b, y) = (\theta_1, y_2^{-1}, \theta_{12})$. L'égalité (3) vient de l'hypothèse de récurrence, (4) de $\det \theta = (\det \theta_1)(\det(\theta_2 - \theta_{21}\theta_1^{-1}\theta_{12}))$ et (5) de la définition 6.7) de la fonction $\Gamma_r(p)$. La puissance de π est la bonne, puisque si $r = m + n$ on a

$$\frac{m(m-1)}{4} + \frac{n(n-1)}{4} + \frac{mn}{2} = \frac{r(r-1)}{4}. \tag{6.9}$$

La Proposition 2 est donc montrée.

VII Lévy processes, reciprocity and the Zolotarev formula.

Proving the existence of an NEF with a given variance function V defined on an open subset M of the linear space E can be a difficult problem. The steps are

1. Finding $\psi : M \to E^*$ such that $(V(m))^{-1} = \psi(m)$ (assuming the necessary condition on V that the bilinear map on E defined by $(u, v) \mapsto V(m)(V(m)u)(v)$ is symmetric in (u, v)).
2. Inverting the map $m \mapsto \psi(m)$ in order to get the differential of the cumulant function k'.
3. Computing the cumulant function k and $L = \exp k$.
4. Checking that L is the Laplace transform of some positive measure.

The hard parts are steps 2 and 4. The Lagrange formula is often helpful for step 2. But the best tool for step 4 is the finding of a probabilistic interpretation.

Let us formalize in a definition the link between two exponential families on $\mathbb{R}$ appearing in (**??**). Suppose that the NEF F_1 on $\mathbb{R}$ with domain of the means M_{F_1} is such that $\tilde{M}_{F_1} = M_{F_1} \cap (0, \infty)$ is not empty. Then the NEF F_2 on $\mathbb{R}$ is called the *reciprocal* NEF of F_1 if $V_2(m) = m^3 V_1(\frac{1}{m})$ for all $m \in \tilde{M}_{F_2}$.

Not all NEF have reciprocal : the NEF generated by a positive stable distribution whose parameter is in $(0, 1)$ has no reciprocal. But suppose that we want to prove the existence of a NEF with variance function $m^3 + m^2$. By translation this is equivalent to the existence of a NEF F_2 with variance function $m(m-1)^2$ which would be the reciprocal of F_1 with variance function $(1-m)^2$ with F_1 concentrated on $(-\infty, 1)$. The NEF F_1 exists, this is nothing but the NEF generated by the Lebesgue measure restricted to $(-\infty, 1)$. Actually we have the following result (Letac-Mora (1990)) :

Theorem 5 : Let $(X(t)_{t \geq 0}$ be a Lévy process with Lévy measure concentrated on the negative line. Let $T(x)$ be the hitting time of $x > 0$. Then the exponential families F_1 and F_2 respectively generated by the distribution of $X(1)$ and the distribution of $T(1)$ (restricted to $(0, \infty)$) are reciprocal. Furthermore the distributions of $X(t)$ (restricted to the positive line) and T_x are related by the following Zolotarev's formula, which indicates the coincidence of two measures on $(0,\infty)^2$

$$xP(X(t) \in dx)dt = tP(T(x) \in dt)dx. \tag{7.10}$$

This magic formula (7.10) has actually been given by Zolotarev (1964). Borovkov (1964) and Dozzi and Vallois (1997) give other proofs. I have learned the above elegant formulation in Bertoin (1999). Let us insist on the fact that (7.10) is not an absolutely continuous measure on $(0,\infty)^2$: only the margins have densities. For instance, if $X(t) = at - bN(t)$, where $a > 0$ and $b > 0$ and $N(t)$ is a Poisson process with intensity λ then the measure (7.10) is concentrated on the lines $x = at - bn$ where $n \in \mathbf{N}$.

As an example, we apply the theorem to $X(t) = t - Y(t)$ where Y is the standard gamma process ($\mathbf{E}(e^{-sY(t)}) = (1+s)^{-t}$). The variance for $X(1)$ being $(1-m)^2$, then the existence of the variance function $m(m-1)^2$ follows. A result similar to Theorem 5 can be obtained with the right continuous random walks in the integers (see Letac-Mora (1990)), providing a relatively explicit generating measure for F_2 (the Ressel Kendall distribution in our example). The same is true for the random walk case, where Lagrange replaces Zolotarev.

However, it is false to think that any reciprocal pair has a similar probabilistic interpretation. Problem :

Why do we have always a kind of Zolotarev formula in case of reciprocity ?

To be more specific, let us say that two measures μ_1 and μ_2 in $\mathcal{M}(\mathbb{R})$ are reciprocal if the sets

$$\tilde{\Theta}(\mu_i) = \{\theta \in \Theta(\mu_i); k_{\mu_i}(\theta) > 0\}$$

are not empty and such that the map $\theta \mapsto -k_{\mu_i}(\theta)$ is a one to one map from $\tilde{\Theta}(\mu_i)$ onto $\tilde{\Theta}(\mu_{3-i})$ whose inverse is $\theta \mapsto -k_{\mu_{3-i}}(\theta)$. Needless to say, under these circumstances, the NEF's $F(\mu_1)$ and $F(\mu_2)$ are reciprocal. A tentative of clarification is offered by the following conjecture (which is even not quite correct, see example 2 below).

Conjecture : Let $\mu(dx)$ and $\nu(dt)$ in $\mathcal{M}(\mathbb{R})$ be reciprocal. Denote by $\lambda(dx)$ and $\eta(dt)$ the measures on $[0,\infty)$ of the form $\sum_{n=0}^{\infty}\delta_{an+b}$ or $\mathbf{1}_{[0,\infty)}(x)dx$ such that $\mathbf{1}_{[0,\infty)}(x)\mu(dx)$ and $\mathbf{1}_{[0,\infty)}(t)\nu(dt)$ are absolutely continuous with respect to $\lambda(dx)$ and $\eta(dt)$ respectively (assuming the existence of $\lambda(dx)$ and $\eta(dt)$). Then the following equality between measures on $\overline{\Lambda}(\nu) \times \overline{\Lambda}(\mu)$ holds :

$$x\mu_t(dx)\eta(dt) = t\nu_x(dt)\lambda(dx). \tag{7.11}$$

Let us give now examples of reciprocity where neither the conditions of Theorem 5 nor the conditions of its right continuous random walks analog are fulfilled.

Example 1 : We take $\mu(dx) = \lambda(dx) = \mathbf{1}_{(0,\infty)}dx$. Thus the Jorgensen set $\Lambda(\mu)$ is $(0,\infty)$, and we have $\mu_t(dx) = \frac{x^{t-1}}{\Gamma(t)}\lambda(dx)$. The reciprocal measure of μ is $\nu(dt) = \frac{1}{\Gamma(t+1)}\eta(dt)$ where $\eta(dt) = \sum_{n=0}^{\infty}\delta_n(dt)$. The Jorgensen set $\Lambda(\nu)$ is $(0,\infty)$, and we have $\nu_x(dt) = \frac{x^t}{\Gamma(t+1)}\eta(dt)$. Clearly (7.11) is satisfied, and the reciprocity is the reciprocity of the Poisson NEF and the exponential distributions NEF with respective variance functions $V_\nu(m) = m$ and $V_\mu(m) = m^2$. We are not in the conditions of application of Theorem 5. The result lacks of a probabilistic explanation.

Example 2 : We take $\mu(dx) = \sum_{n=0}^{\infty} \frac{1}{n!}\delta_{n-1}(dx)$. Thus $M_\mu = (-1, \infty)$ and $V_\mu(m) = m + 1$: this is a shifted Poisson family. The Jorgensen set is $(0, \infty)$ and

$$\mu_t(dx) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\delta_{n-t}(dx).$$

The reciprocal family has therefore a variance function equal to $V_\nu(m) = m^2(1+m)$ which is a Ressel Kendall family.

One can prove (see Letac-Mora (1990)) that the reciprocal measure ν of μ has a Jorgensen set equal to $(0, \infty)$ and that the measures $\nu_x(dt)$ are explicitly given by

$$\nu_x(dt) = \frac{xt^{x+t-1}}{\Gamma(x+t+1)}\eta(dt)$$

where $\eta(dt) = \mathbf{1}_{(0,\infty)}(t)dt$. In this case the conjecture is not quite satisfied, since the reference measure $\lambda(dx)$ is the restriction to the positive line of

$$\sum_{a \in \mathbf{N}-t} \delta_a(dx),$$

which depends a little bit of t. Up to this (7.11) is satisfied.

Example 3 : We take μ as the distribution of the difference of two independent Poisson random variables with means 1/2. Thus $M_\mu = \mathbb{R}$ and $V_\mu(m) = (m^2 + 1)^{1/2}$, $\lambda(dx) = \sum_{n=0}^{\infty} \delta_n(dx)$, μ is infinitely divisible and since we have

$$e^{t \cosh \theta - t} = \sum_{n \in \mathbf{Z}} \mu_t(n)e^{n\theta},$$

thus $\mu_t(n) = e^{-t}I_{|n|}(t)$ where

$$I_x(t) = \sum_{n=0}^{\infty} \frac{1}{n!\Gamma(n+x+1} \left(\frac{t}{2}\right)^{2n+x}.$$

The reciprocal family does exist and has variance function $V_\nu(m) = m^2(m^2 + 1)^{1/2}$. It is generated by the reciprocal measure $\nu(dt) = e^{-t}\frac{1}{t}I_1(t)\eta(dt)$ where $\eta(dt) = \mathbf{1}_{(0,\infty}(t)dt$. See Feller (1966) pages 414, formula (3.8) and page 427, example (d). This is also infinitely divisible and (7.11) holds.

References

BAR-LEV, S., ENIS, P., LETAC, G. (1994). Sampling models which admit a given general exponential family as a conjugate family of priors. *Ann. Statist.* **22**, 1555-86.

BERTOIN, J. (1999). *Subordinators : Examples and Applications.* Ecole d'été de probabilités de Saint Flour, XXVII. Lecture Notes in Math. **1727**, Springer.

BOROKOV, A. A. (1964). On the first passage time for one class of processes with independent increments. *Theor. Prob. Appl.* **10**, 331-334.

CASALIS, M. (1991). Les familles exponentielles à variance quadratique homogène sont des lois de Wishart sur un cône symétrique. *C. R. Acad. Sci. Paris Sér. I Math.* **312**, 537-540.

CASALIS, M. (1996). The $2d+4$ simple quadratic natural exponential families on $\mathbb{R}^d$. *Ann. Statist.* **24**, 1828-1854.

DOZZI, M., VALLOIS, P. (1997). Level crossing for certain processes without positive jumps. *Bull. Sci. math.* **121**, 355-376.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications.* Wiley, New York.

HASSAÏRI, A. (1992). La classification des familles exponentielles sur $\mathbb{R}^n$ par l'action du groupe linéaire sur $\mathbb{R}^n$. *C. R. Acad. Sci. Paris Sér. I Math.* **315**, 207-210.

HASSAÏRI, A. (1994). Classification des familles exponentielles dans $\mathbb{R}^d$ de variance cubique du type Mora. *Doctorat de l'université Paul Sabatier, Toulouse.*

LETAC, G., MORA, M. (1990). Natural exponential families with cubic variance functions. *Ann. Statist.* **18**, 1-37.

LETAC, G. (1992). *Lectures on Natural Exponential Families and their Variance Functions. Monografias de Matemática* **50**, IMPA, Rio de Janeiro.

MORRIS, C. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10**, 65-80.

ZOLOTAREV, V. M. (1964). The first passage time of a level and the behavior at infinity for one class of processes with independent increments. *Theor. Prob. Appl.* **9**, 653-664.