FINITELY GENERATED CUMULANTS

Giovanni Pistone and Henry P. Wynn

Politecnico di Torino and University of Warwick

Abstract: Computations with cumulants are becoming easier through the use of computer algebra but there remains a difficulty with the finiteness of the computations because all distributions except the normal have an infinite number of non-zero cumulants. One is led therefore to replacing finiteness of computations by "finitely generated" in the sense of recurrence relationships. In fact it turns out that there is a natural definition in terms of the exponential model which is that the first and second derivative of the cumulant generating function, K, lie on a polynomial variety. This generalises recent polynomial conditions on variance functions. This is satisfied by many examples and has applications to, for example, exact expressions for variance functions and saddle-point approximations.

Key words and phrases: Computer algebra, cumulants, exponential models.

1. Introduction

It is perhaps best to introduce this paper by describing briefly the route by which the authors came to the definition of finite generation of cumulants. The starting point was the recognition that in the understanding of the propagation of randomness through systems, cumulants may be useful. The cumulants of the output Y of a system may be computed, for some systems, directly from the cumulants of the input X. The theory of McCullagh (1987) allows this to be done if the function $y = f(x)$ relating the input to the output is polynomial, and where both X and Y are multivariate. We mention also the survey paper of Gutiérrez-Peña and Smith (1997), where the relation to conjugate prior distributions is discussed.

The authors are aware of the widespread use of cumulants as "higher order statistics" in signal processing. It was these topics which suggested that there may be a computational theory or, at least with the availability of fast computational algebra packages, there should be ways of setting up cumulant calculations in an attractive fashion. This had already been recognized by researchers in stochastic geometry, see Kendall (1993).

The first difficulty in trying to set up such a methodology is that apart from the normal distribution all other distributions have an infinite number of non-zero cumulants (Marcinkievicz Theorem, see Feller (1966)). This means that any attempt to produce a closed algebraic system with a finite number of operations

by setting cumulants to zero above a finite order is fallacious. For non-normal variables one is left with expanding the computations as the order gets higher, as in the McCullagh theory. And yet there are many cases where the cumulants are very simple to compute, such as the Poisson distribution where they are all equal. This points towards using recurrence relationships, replacing the idea of "finiteness" by "finitely generated". After considerable exploration even this does not seem the most appropriate environment, although it is quite close to the definition finally adopted.

Two other strands of research have had an impact. First the work on the exponential models, and particularly that of Letac (1992), points to the pivotal role of the cumulant generating function relationship in the definition of variance function. The use of the exponential model in areas such as saddle-point approximation is already established, see the books by Barndorff-Nielsen and Cox (1989, 1994). Second, the authors recent work (Pistone and Wynn (1996)) on the use of Gröbner bases in identification in experimental design points to elimination theory as a useful tool for computations.

2. The Exponential Model

The question we address in this section has two parts: the relationship of the cumulant function K to the exponential family and the consequences for computations with cumulants.

Provided certain regularity conditions hold, an "exponential model", or equivalently a natural exponential family of distributions, can be associated with any (possibly vector) random variable whose Laplace transform is defined in a neighborhood of 0. This important idea is due to Khinchin who introduced it in the context of Statistical Mechanics (Khinchin (1949)), and was developed by a number of authors (see the papers by Morris (1982, 1983) for the first applications to characterization and Casalis (1996)), but particularly over a number of years by Letac (1992). This is somewhat different from saying that the random variable comes from, or lies in, an exponential family. Relevant references on exponential models are Barndorff-Nielsen (1978), Johansen (1979), Brown (1986) and Letac (1992). See also Pistone and Sempi (1995), Pistone and Rogantin (1998) and Gibilisco and Pistone (1998) for non-parametric generalizations.

2.1. Basic properties

The following definitions and theorems are all well known from the references already given. We have collected them here for ease of reference. If no confusion is possible, we denote by ab the scalar product of vectors a and b. Otherwise we use $a \cdot b$.

Definition 1. Let X be a random vector in $\boldsymbol{R}^m$. Denote by D_X the interior of the convex set $\{t \in \boldsymbol{R}^m : E[e^{tX}] < +\infty\}$. If $D_X \neq \emptyset$ then the moment (generating) function M_X and cumulant (generating) function K_X of X are the functions defined for each $t \in D_X$ by $M_X(t) = E[e^{tX}]$, $K_X(t) = \log M_X(t)$.

Theorem 2. *Assume that the random vector $X = (X_1, \ldots, X_m)$ has generating functions M_X and K_X with domain D_X. Then:*
1. *The moment function M_X and the cumulant function K_X are convex. If X is not a constant they are strictly convex;*
2. *The moment function M_X and the cumulant function K_X are analytic in D_X. The derivatives of the moment function are given by the equations*

$$\frac{\partial^{n_1 + \ldots + n_m}}{\partial t_1^{n_1} \ldots \partial t_m^{n_m}} M_X(t) = E[X_1^{n_1} \ldots X_m^{n_m} e^{tX}];$$

3. *The first derivative of the cumulant function*

$$K_X'(t) = (\frac{\partial}{\partial t_1} K_X(t), \ldots, \frac{\partial}{\partial t_m} K_X(t))$$

is a 1-to-1 mapping from D_X to $\boldsymbol{R}^m$;
4. *If the generating functions are defined in a neighborhood of 0, then the random vector X has finite moments of all orders and the raw moments are given by*

$$\mu_{n_1 \ldots n_m}(X) = \left.\frac{\partial^{n_1 + \cdots + n_m}}{\partial t_1^{n_1} \ldots \partial t_m^{n_m}} M_X(t)\right|_{t=0}.$$

Definition 3. If the domain D_X of the generating functions of the random vector X contains 0, we will say that X belongs to the class $\mathcal{E}$ of exponentially integrable random vectors; in such a case the coefficients of the Taylor series of the cumulant function at 0 are called cumulants:

$$\kappa_{n_1 \ldots n_m}(X) = \left.\frac{\partial^{n_1 + \cdots + n_m}}{\partial t_1^{n_1} \ldots \partial t_m^{n_m}} K_X(t)\right|_{t=0}.$$

Definition 4. Let X be a random vector of class $\mathcal{E}$ and let D_X be the domain of the generating functions. Then the equation

$$p(x; \theta) = e^{\theta x - K_X(\theta)}, \quad \theta \in D_X, \tag{1}$$

defines an exponential model with respect to the distribution F_X of the random vector X. Such a model is called the natural exponential model associated to X. Such a model can be parameterized by the mean parameter $\eta = K_X'(\theta)$ (see Theorem 2, Part 3).

Notice that all exponential models are exactly of this form when parameterised by the natural exponential parameter. Moreover the random variable X is the sufficient statistics of the exponential model and $U = X - E[X]$ is the score statistic at $\theta = 0$. The exponential model associated to the random variable X can be sought as an exponential model of probability distributions, starting at the distribution F_X of X, in the "direction" U.

The results on the generating functions take an extended form when considered with respect to the exponential model. We denote by E_θ the expectation with respect to the θ-distribution in the natural exponential model of the random vector X.

Again we try to avoid long notation in the multivariate case, by writing for a multi-index $(n_1, \ldots, n_m)$:

$$n! = n_1! \cdots n_m!,$$
$$t^n = t_1^{n_1} \cdots t_m^{n_m},$$
$$\frac{d^n}{dt^n} f(t) \text{ or } f^{(n)}(t) = \frac{\partial^{n_1 + \cdots + n_m}}{\partial t_1^{n_1} \ldots \partial t_m^{n_m}} f(t_1, \ldots, t_m).$$

Theorem 5. *The random vector X is exponentially integrable with respect to all E_θ, $\theta \in D_X$, and*

1. *The random vector $X = (X_1, \ldots, X_m)$ has finite θ-moments of all orders and the raw moments are given by*

$$E_\theta[X^n] = M_X(\theta)^{-1} \frac{d^n}{d\theta^n} M_X(\theta);$$

2. *The coefficients of the Taylor series of K_X at θ are the cumulants of X with respect to θ:*

$$K_X(\theta + t) - K_X(\theta) = \sum_{n \geq 1} K_X^{(n)}(\theta) \frac{t^n}{n!}.$$

Proof. The generating functions of X with respect to θ are

$$M_{X,\theta}(t) = E_\theta[e^{tX}] = \int e^{\theta x - K_X(\theta)} e^{tx} \, dF_X(x) = \frac{M_X(t + \theta)}{M_X(\theta)},$$
$$K_{X,\theta}(t) = \log M_{X,\theta}(t) = K_X(t + \theta) - K_X(\theta).$$

The moment function M_X and the cumulant function K_X are connected by the relation $M_X = e^{K_X}$. By successive derivation and substitution it is easy to prove the following proposition. We do not consider the (straightforward) derivation of the analogous multivariate formulae.

Proposition 6. *For a real random variable X of the class $\mathcal{E}$ and in a suitable neighborhood of 0:*

1. *For all $n \geq 0$,*

$$M_X^{(n+1)}(t) = \sum_{h=0}^{n} \binom{n}{h} K_X^{(h+1)}(t) M_X^{(n-h)}(t);$$

2. *For all $n \geq 1$,*

$$M_X^{(n)}(t) = M_X(t) G_n(K_X'(t), \ldots, K_X^{(n)}(t)),$$

where the polynomials $G_n(\kappa_1, \ldots, \kappa_n)$ are defined by

$$G_1(\kappa_1) = \kappa_1,$$
$$G_{n+1}(\kappa_1, \ldots, \kappa_{n+1}) = \kappa_1 G_n(\kappa_1, \ldots, \kappa_n)$$
$$+ \sum_{i=1}^{n} \frac{\partial}{\partial \kappa_i} G_n(\kappa_1, \ldots, \kappa_n) \kappa_{i+1};$$

3. *For all $n \geq 1$*

$$M_X^{(n)}(t) K_X^{(n)}(t) = H_n(M_X(t), M_X'(t), \ldots, M_X^{(n)}(t)),$$

where the homogeneous polynomials $H_n(\mu_0, \ldots, \mu_n)$ are defined by

$$H_1(\mu_0, \mu_1) = \mu_1,$$
$$H_{n+1}(\mu_0, \ldots, \mu_{n+1}) = \mu_0 \sum_{i=0}^{n} \frac{\partial}{\partial \mu_i} H_n(\mu_0, \ldots, \mu_n) \mu_{i+1}$$
$$- n \mu_1 H_n(\mu_0, , \ldots, \mu_n). \tag{2}$$

Proof. Use the basic relation $M_X' = M_X K_X'$. 1. is Leibnitz's formula; 2. and 3. are verified by recurrence.

2.2. Variance function

The construction of the natural exponential family allows a clean definition of the variance function related to the original random variable X.

Definition 7. Let Ψ_X be the inverse of the gradient of the cumulant function K_X, see Theorem 2, Part 3. The *variance function* of the exponential model (1) is defined by

$$V_X(\eta) = K_X''[\Psi_X(\eta)], \tag{3}$$

where K_X'' denotes the Hessian matrix of the cumulant function.

The value of the variance function $V_X(\eta)$ is the variance of the random vector X with respect to the unique distribution of the exponential model of X such that the expectation of X is η.

A number of authors have classified distributions for which the variance function $V_X(\cdot)$ has a particular form. For example the class when $V_X(\cdot)$ is quadratic was classified by Morris (1982) into six types: normal, Poisson, binomial, negative binomial, gamma and hyperbolic cosine. The cubic case has been studied by Mora (1986) and by Letac and Mora (1990). In the next section we propose and develop a new definition which is particularly suited to computational algebra.

Another place where cumulant generating functions are used is in classical results on the large deviation theory for exponential families. If K_X is the cumulant function of the random vector X, with domain D_X, then its *conjugate function* is the convex function $H_X(\eta) = \sup_\theta \{\theta\eta - K_X(\theta)\}$. The value of the conjugate function at θ can be computed by solving the *likelihood equations* $\eta = K_X'(\theta)$, with $\theta = \Psi_X(\eta)$, and substituting to get $H_X(\eta) = \eta\Psi_X(\eta) - K_X[\Psi_X(\eta)]$. It follows that $H_X' = \Psi_X$, and

$$K_X'\left[H_X'(\eta)\right] = \eta, \tag{4}$$

$$K_X''\left[H_X'(\eta)\right] H_X''(\eta) = 1. \tag{5}$$

3. Finitely Generated Cumulants

If the variance function $V_X(\cdot)$ in (3) is polynomial in each component, then differentiation of $K_X''(t) = V(K_X'(t))$ with respect to the i-component of $t = (t_1, \ldots, t_n)$, followed by substituting $t = t_0$, will give all the coefficients of the Taylor development of $K_X(t)$ at t_0, that is all the cumulants of X with respect to the natural exponential model of X at t_0, by means of algebraic computations.

Another possibility is to follow Morris (1982) and observe that the cumulants can be expressed as a function $C_r(\eta)$ of the mean parameters η, so that

$$C_{r+1}\left[K'(t)\right] = \frac{d}{dt}C_r\left[K'(t)\right] = C_r'\left[K'(t)\right]K''(t) = C_r'\left[K'(t)\right]V\left[K'(t)\right]$$

and

$$C_{r+1}(\eta) = C_r'(\eta)V(\eta), \quad r > 1.$$

This, together with $C_1(\eta) = \eta$, implies, in the case of a polynomial variance function, that the cumulants as functions of the mean parameters are polynomials.

The main contribution of this paper is to allow an implicit polynomial dependence.

The algebraic background of our theory is best described using the commutative ring theory as described for example in Cox, Little and O'Shea (1992). We refer to it for the notions of number field, polynomial ring, ideal. If k is a numeric field, and $x_1, \ldots, x_d$ are indeterminates, we denote by $k[x_1, \ldots, x_d]$ the polynomial ring, e.g. the set of all polynomials in the indeterminates $x_1, \ldots, x_d$, with

coefficients in k, with the usual algebraic operations. If polynomials $g_1, \ldots, g_n$ are given in the polynomial ring, we denote by $\mathrm{Ideal}(g_1, \ldots, g_n)$ the ideal generated by the given polynomials, e.g. the set of all polynomials of the form $\sum_i h_i g_i$ for arbitrary polynomials $h_1, \ldots, h_n$ in the same polynomial ring. This roughly coincides with the set of all polynomials which are 0 on the variety in k^d defined by the equations $g_1 = 0, \ldots, g_n = 0$. Examples of number fields we use are the rationals $\boldsymbol{Q}$ or fields of the type $\boldsymbol{Q}(\sqrt{2})$ which consists of all numbers of the type $a + b\sqrt{2}$ with $a, b \in \boldsymbol{Q}$.

Definition 8. Let k be a number field. A polynomial $G \in k[\eta, \gamma]$ is a *generating polynomial* at the point $(\bar{\eta}, \bar{\gamma}) \in k^2$ if $G(\bar{\eta}, \bar{\gamma}) = 0$ and $\frac{\partial}{\partial \gamma} G(\bar{\eta}, \bar{\gamma}) \neq 0$. The condition in the previous definition implies an implicit function property, but it is of an algebraic type. In a general way it can be checked as follows. We want to check that $G(\eta, \gamma) \in \mathrm{Ideal}\,(\eta - \bar{\eta}, \gamma - \bar{\gamma})$, but $\frac{\partial}{\partial \gamma} G(\eta, \gamma) \notin \mathrm{Ideal}\,(\eta - \bar{\eta}, \gamma - \bar{\gamma})$. This can be done in a generic way by computing all the $(\bar{\eta}, \bar{\gamma})$'s with respect to which $G(\eta, \gamma)$ is a generating polynomial.

In the following we will avoid explicit mention of the underlying number field, which will be clear from the specific case. See also Section 4 for more details and references on polynomial rings theory.

Definition 9. The cumulants of X are called *finitely generated* if there exist polynomials

$$F_{hk}(\eta_i : i = 1, \ldots, m; \gamma_{ij} : i \leq j = 1, \ldots, m), \quad h \leq k = 1, \ldots, m,$$

such that the corresponding system of equations can be uniquely solved for $\gamma = (\gamma_{ij})_{1 \leq h \leq k \leq m}$ as a function of $\eta = (\eta_i)_{1 \leq i \leq m}$, around the point $\eta_0 = K'_X(0)$, $\gamma_0 = K''_X(0)$, and the equations

$$F_{hk}(K'_X(t), K''_X(t)) = 0, \quad h \leq k = 1, \ldots, m, \tag{6}$$

hold in a neighborhood of 0. The polynomials $F = (F_{hk})_{h \leq k = 1, \ldots, m}$ are called *generating polynomials of X* and equation (6), generating equations.

We shall refer to the property in Definition 9 as the FGC (finitely generated cumulant) property. The existence of the variance function $V_X(\cdot)$ together with the definition will ensure that at most the pair $(\eta, \gamma) = (\eta, V_X(\eta))$ is a solution to each $F_{hk}(\eta, \gamma) = 0$, $1 \leq h \leq k \leq m$, in a suitable neighborhood of $(K''_X(0), K'_X(0))$.

Moreover differentiation of $F(K'_X(t), K''_X(t))$ with respect to $(t_1, \ldots, t_n)$ and putting all $t_i = t_{0,i}$ will again give a finite algorithm for generation of the coefficients of the Taylor development of $K(t)$ at t_0.

The present paper will mostly consider the one-dimensional case. The multi-dimensional case will be considered only for vectors with independent components or generated by linear transformations of such a case. For the case of multidimensional variance functions we refer to the work by Letac (1992).

Proposition 10. *Let the generating polynomial $F(\eta, \gamma)$ of the random variable X be of degree d in γ. Let H_X denote the conjugate of the cumulant function K_X. Then the conjugate function satisfies the polynomial non-autonomous differential equation $G\left(\eta, H''_X(\eta)\right) = 0$, where $G(\eta, \psi) = \psi^d F(\eta, \psi^{-1})$.*

Proof. Substitute $\theta = H'_X(\eta)$ in the generating equation, and use (4) and (5).

Proposition 11. *Let the random variable X have finitely generated cumulants with generating polynomial $F(\kappa_1, \kappa_2)$. The equations obtained by deriving $n \geq 1$ times the generating equation $F(K'_X(t), K''_X(t)) = 0$ are of the form*

$$L_n\left(K'_X(t), K''_X(t), \ldots, K^{(n+1)}(t)\right) + F_2(K'_X(t), K''_X(t))K_X^{(n+2)}(t) = 0, \qquad (7)$$

where

$$F_2(\eta, \gamma) = \frac{\partial}{\partial \gamma} F(\eta, \gamma),$$

and the polynomials $L_n(\kappa_1, \ldots, \kappa_{n+1})$ are defined by

$$L_1(\kappa_1, \kappa_2) = \frac{\partial}{\partial \kappa_1} F(\kappa_1, \kappa_2),$$

$$L_{n+1}(\kappa_1, \ldots, \kappa_{n+2}) = \sum_{i=1}^{n+1} \frac{\partial}{\partial \kappa_i} L_n(\kappa_1, \ldots, \kappa_{n+1})\kappa_{i+1}$$
$$+ \left(\frac{\partial}{\partial \kappa_1} F_2(\kappa_1, \kappa_2)\kappa_2 + \frac{\partial}{\partial \kappa_2} F_2(\kappa_1, \kappa_2)\kappa_3\right)\kappa_{n+1}.$$

In particular, the equations at (7) are first order and uniquely solvable in the higher derivative $K_X^{(n+2)}$ around the point $\left(K'_X(0), K''_X(0), \ldots, K_X^{(n+2)}(0)\right)$.

Proof. Remember that the partial derivative $F_2(\eta, \gamma)$ is non-zero at the point $(K'_X(0), K''_X(0))$ because of the FGC condition. Then all points can be checked directly on the equations.

For random variables X with finitely generated cumulants successive differentiation of F followed by setting $t = t_0$ will generate a family of polynomial equations for the cumulants. These can then be added to those arising from the well known relationship between moments and cumulants and its generalization to the multivariate case to establish values of moments and cumulants all expressed in terms of the vector of mean $K'_X(t)$.

The following seems to be an unsolved problem in statistics: find those polynomials $F(\cdot, \cdot)$ such that $F(K', K'') = 0$ for the cumulant distribution function K of some random variable X. Typically this problem is stated in the more restrictive fashion in terms of the quadratic or cubic nature, and so on, of the variance function. We believe that casting the problem in the implicit form $F(K', K'') = 0$ may be preferable.

3.1. Polynomial transformation

A useful application of the FGC condition is when we take polynomial functions of the random variable X, say $Y = h(X)$, where $Y = (Y_1, \ldots, Y_m)$, see McCullagh (1984, 1987). Remember that a polynomial function of a random variable of class $\mathcal{E}$ has all moments but it is not in general class $\mathcal{E}$, so that the cumulant theory does not apply in all cases, but formally as a way to compute recursively the moments.

Example 12. Consider X^2 with X exponentially distributed. In such a case the cumulants are *formally* defined as polynomial transformations of the moments by the equations

$$\kappa_n(Y) = H_m(1, \mu_1(Y), \ldots, \mu_n(Y)),$$
$$= H_m(1, \mu_2(X), \ldots, \mu_{2n}(X)),$$

where the polynomials H_n are computed using (2). The exponential family generated by Y does not exist because $e^{tY} = e^{tX^2}$ is never integrable for $t \neq 0$.

Since all the moments of Y can be computed from those of X and all those of X computed in a finite algorithm if X has finitely generated cumulants the moments of Y can be also computed in a finite algorithm. We can express this as two steps.

Definition 13. The (formal) cumulants of X are said to be weakly finitely generated if there is an infinite sequence of polynomials F_n, $n = 1, 2, \ldots$ such that each F_n has n indeterminates $F_n(\kappa_1(X), \kappa_2(X), \ldots, \kappa_n(X)) = 0$ for all $n \geq 1$, and each equation is uniquely solvable in the formal cumulant of maximum order.

If the cumulants of X are finitely generated then they are weakly finitely generated, and the sequence F_n is obtained by derivation from the basic ones.

Proposition 14. *If the cumulants of a random variable X are finitely (or the formal cumulants are weakly finitely) generated then the cumulants of any polynomial function of X, $Y = h(X)$, are weakly finitely generated.*

Proof. First use $Y = h(X)$ to derive the moments of Y as polynomial functions of the moments of X. Then use Proposition 6 to express the formal cumulants

of Y as functions of the formal cumulants of X. Finally use the FGC property and its derivatives, as in Proposition 11, or its weak counterpart, to eliminate the cumulants of X.

3.2. Examples

We pause to give some simple examples of densities with the FGC property.

Example 15. Among the first examples are the distributions with *quadratic variance function*, as discussed by Morris in (1982), Table 1, reproduced here as Table 1 in our notation.

Table 1. Quadratic variance.

Distribution	Parameters	Generating polynomial
Normal	λ, σ^2	$K'' - \sigma^2$
Poisson	λ	$K'' - K'$
Gamma	r, λ	$rK'' - (K')^2$
Binomial	r, p	$rK'' - K'(r - K')$
Negative Binomial	r, p	$rK'' - K'(r + K')$
GHS	r, λ	$rK'' - (K')^2 - r^2$

More elaborate examples are not easily computed by hand and need some algebraic elimination theory, to be discussed in the next section. We give here only one example in which the variance function is not a polynomial.

Example 16. The *Laplace* (*double exponential*) *density* with parameter 1 has cumulant function $K(t) = -\log(1 - t^2)$. Then the first and second derivatives are

$$K'(t) = \frac{2t}{1 - t^2}, \tag{8}$$

$$K''(t) = 2\frac{1 + t^2}{(1 - t^2)^2}. \tag{9}$$

From these equations it follows that

$$(K'')^2 - 2(1 + (K')^2)K'' + (K')^2 + (K')^4 = 0, \tag{10}$$

as can be checked by substitution.

3.3. First properties of FGC

The notion of FGC would not be useful without some stability with respect to the usual statistical transformations. The following proposition gives the first results. More properties are discussed later.

Proposition 17. *The FGC property is stable for*

1. *joining independent components, in particular sampling;*
2. *invertible linear transformations;*
3. *convolutions and de-convolution of the same distribution.*

Proof.

1. Let $X_1, \ldots, X_n$ be independent with cumulant functions $K_1, \ldots, K_n$ and generating polynomials $F_1(\eta_1, \gamma_1), \ldots, F_n(\eta_n, \gamma_n)$. The random vector $X = (X_1, \ldots, X_n)$ has cumulant function $K(t_1, \ldots, t_n) = K_1(t_1) + \cdots + K_n(t_n)$, with derivatives

$$\frac{\partial}{\partial t_i} K(t_1, \ldots, t_n) = K_1'(t_i),$$

$$\frac{\partial^2}{\partial t_i^2} K(t_1, \ldots, t_n) = K_1''(t_i),$$

$$\frac{\partial^2}{\partial t_i \partial t_j} K(t_1, \ldots, t_n) = 0, \quad i \neq j.$$

The generating polynomials of X are given by

$$F_{ii}((\eta_i)_{i=1,\ldots,n}, (\gamma_{ij})_{i,j=1,\ldots,n}) = F_i(\eta_i, \gamma_{ii}),$$
$$F_{ij}((\eta_i)_{i=1,\ldots,n}, (\gamma_{ij})_{i,j=1,\ldots,n}) = 0, \quad i \neq j.$$

2. Let $K(t)$ be the cumulant function of the random vector X with components $X_1, \ldots, X_n$, and let

$$F(\eta, \gamma) = [F_{ij}((\eta_i)_{i=1,\ldots,n}, (\gamma_{ij})_{i,j=1,\ldots,n})]_{i,j=1,\ldots,n}$$

be the (matrix of the) generating polynomials. If A is a non-degenerate $n \times n$ matrix, then the cumulant function of the random vector $Y = AX$ is $K_Y(t) = K(Bt)$, where $B = $ transpose of A. It follows that $K_Y'(t) = BK'(Bt)$, $K_Y''(t) = BK''(Bt)A$, and the generating polynomials of Y are $F(B^{-1}\eta, B^{-1}\gamma A^{-1})$.

3. The cumulant function of the sum of independent and identically distributed $X_1, \ldots, X_n$ is $K_{X_1+\cdots+X_N}(t) = nK_{X_1}(t)$. If $F(\eta, \gamma)$ is the generating polynomial for X_1 then $F\left(\frac{1}{n}\eta, \frac{1}{n}\gamma\right)$ is the generating polynomial for the sum. Conversely if the nth-fold convolution has the FGC property so does the original distribution.

When a class of random variables has the same exponential model then if one member has the FGC property they all do. This is because the cumulants are obtained by a shift in θ, $K_Y(\theta) = K_X(\theta + c)$, and this shift carries over to $F(K', K'') = 0$.

In the next section we give some cases which lie outside the cases mentioned in the last section. They point to a rich theory connecting distributions with polynomial ideals.

4. The Use of Elimination Theory

We first discuss the idea of elimination in a little more detail. Algebraic elimination can be performed with various computer algebra systems. Here we present the elimination algorithms based on Gröbner bases included in Maple, see Char, Geddes, Gonnet, Leong, Monogan and Watt (1991). Other more specialized algorithms exist, for example in the software CoCoA, see Capani, Niesi and Robbiano (1995).

Let k be a number field. Given the polynomial ideal

$$I = \text{Ideal} \left(f_1(x_1, \ldots, x_m), \ldots, f_s(x_1, \ldots, x_m) \right),$$

we attempt to eliminate the indeterminates $x_1, \ldots, x_k$ from the equations of the corresponding variety

$$\begin{cases} f_1(x_1, \ldots, x_m) = 0, \\ \qquad\qquad \vdots \\ f_s(x_1, \ldots, x_m) = 0, \end{cases}$$

leaving only equations in $x_{k+1}, \ldots, x_n$. Formally the elimination ideal is the ideal of $k[x_{k+1}, \ldots, x_n]$ defined by

$$I_k = I \cap k[x_{k+1}, \ldots, x_n].$$

The elimination theorem (see Cox, Little and O'Shea (1992), Chapter 3, Theorem 2) allows us to work with the Gröbner basis and `plex` ordering with $x_1 > \cdots > x_n$. Thus if G is a `plex`-Gröbner basis then

$$G_k = G \cap k[x_{k+1}, \ldots, x_n]$$

is a Gröbner basis for the k-elimination ideal above.

This means that as we move down the members of the Gröbner basis the remaining $n - k$ equations contain only the remaining $n - k$ variables. They may not appear in easily solvable form, but at least they are on their own.

We are interested here with the situation in which $K'(\theta)$ and $K''(\theta)$ are expressed implicitly in some variable t. We eliminate t to obtain $F(K'(\theta), K''(\theta)) = 0$. First it is necessary to understand a little more elimination theory. Write, with $f_1, \ldots, f_n$ polynomials

$$\begin{cases} x_1 = f_1(t_1, \ldots, t_r), \\ \qquad\qquad \vdots \\ x_m = f_m(t_1, \ldots, t_r). \end{cases}$$

We first write these in variety form: $x_1 - f_1 = \cdots = x_n - f_n = 0$, giving the ideal of $k[x_1, \ldots, x_m, t_1, \ldots, t_r]$: $\text{Ideal}(x_1 - f_1, \ldots, x_n - f_n)$. Then using `plex`

with $t_1 > \cdots > t_r > x_1 > \cdots > x_m$ we can eliminate $t_1, \ldots, t_r$ using the above elimination theory. We obtain the *smallest* variety in $k[x_1, \ldots, x_m]$ containing the variety projection of the variety

$$\begin{cases} x_1 - f_1 = 0, \\ \qquad \vdots \\ x_n - f_n = 0. \end{cases}$$

on the coordinates $x_1, \ldots, x_m$.

A very useful extension of this idea will be used shortly. This is the case when

$$x_j = \frac{p_j(t_1, \ldots, t_m)}{q_j(t_1, \ldots, t_m)}, \quad j = 1, \ldots, n,$$

where p_j, q_j are polynomials, that is the x_j are rational functions. Now introduce an additional variable y and write

$$\begin{cases} q_1 x_1 - p_1 = 0, \\ \qquad \vdots \\ q_n x_n - p_n = 0, \\ q_1 \cdots q_n y - 1 = 0. \end{cases}$$

The last variety prevents $q_j = 0$ for all $j = 1, \ldots, n$. We then eliminate with `plex` and the order $y > t_1 > \cdots > t_m > x_1 > \cdots > x_n$. We obtain the smallest variety in $k[x_1, \ldots, x_n]$ containing the solutions to the original problem (except $q_1 \ldots q_n = 0$).

Example 18. Consider as an example the *discrete distribution* with mass $1/3$ on $[0, 1, 2]$. Then

$$M(\theta) = \frac{1}{3}(1 + e^\theta + e^{2\theta}),$$
$$K(\theta) = -\log 3 + \log(1 + e^\theta + e^{2\theta}),$$
$$K'(\theta) = \frac{e^\theta + 2e^{2\theta}}{(1 + e^\theta + e^{2\theta})},$$
$$K_X''(\theta) = \frac{e^\theta + 4e^{2\theta} + e^{3\theta}}{(1 + e^\theta + e^{2\theta})^2}.$$

Now replace e^θ by t and we have an implicit relation between K' and K'' considered now simply as variables:

$$\kappa_1 = \frac{t + 2t^2}{1 + t + t^2},$$
$$\kappa_2 = \frac{t + 4t^2 + t^3}{(1 + t + t^2)^2}.$$

Using the Gröbner suite on Maple the command is

$$\mathrm{gbasis}([R1, R2, R3], [t, y, k1, k2], \texttt{plex}),$$

where

$$R1 := k_1(1 + t + t^2) - t - 2t^2,$$
$$R2 := k_2(1 + t + t^2)^2 - t - 4t^2 - t^3,$$
$$R3 := (1 + t + t^2)^3 y - 1.$$

Elimination gives the generating polynomial

$$3(K')^4 + 2K' - 2K'' + 11(K')^2 - 12K'K''$$
$$-12(K')^3 + 6(K')^2(K'') + 3(K'')^2.$$

The smaller solution of the generating equation for K'' in terms of K' gives the variance function

$$K'' = \frac{1}{3} + 2K' - (K')^2 - \frac{1}{3}\sqrt{1 + 6K' - 3(K')^2}.$$

This is the solution which has the value $K''(0) = 2/3$ at the point $K'(0) = 1$. Note that for $R3$ it would have been enough to take $(1 + t + t^2)y$, since the same term appears in $R2$ and $R3$.

By extending this example to arbitrary distributions on a finite number of variables we obtain the following:

Proposition 19. *Every discrete distribution supported on an equally spaced set of reals has the FGC property.*

Proof. Without loss of generality, we assume X is distributed on the integers $\{0, \ldots, n-1\}$ with probabilities $\{p_0, \ldots, p_{n-1}\}$. The moment function and cumulant function of X are

$$M(t) = p_0 + p_1 e^t + \cdots + p_{n-1} e^{(n-1)t}, \tag{11}$$
$$K(t) = \log(p_0 + p_1 e^t + \cdots + p_{n-1} e^{(n-1)t}), \tag{12}$$

and the derivatives of the moment function are

$$M'(t) = p_1 e^t + 2p_2 e^{2t} + \cdots + (n-1)p_{n-1} e^{(n-1)t}, \tag{13}$$
$$M''(t) = p_1 e^t + + 4p_2 e^{2t} + \cdots + (n-1)^2 p_{n-1} e^{(n-1)t}. \tag{14}$$

Form the basic relationships

$$M(t)K'(t) = M'(t),$$
$$M^2(t)K''(t) = M(t)M''(t) - (M'(t))^2,$$

and introduce the new indeterminates $\mu = M(t)$, $\kappa_1 = K'(t)$, $\kappa_2 = K''(t)$, $\zeta = e^t$. The previous system of Equations (11, 13, 14) becomes

$$
\begin{cases}
\mu = p_1\zeta + p_2\zeta^2 + \cdots + p_{n-1}\zeta^{(n-1)}, \\
\mu\kappa_1 = p_1\zeta + 2p_2\zeta^2 + \cdots + (n-1)p_{n-1}\zeta^{(n-1)}, \\
\mu^2\kappa_2 = \mu(p_1\zeta + 4p_2\zeta^2 + \cdots + (n-1)^2 p_{n-1}\zeta^{(n-1)}) \\
\qquad\quad - (p_1\zeta + 2p_2\zeta^2 + \cdots + (n-1)p_{n-1}\zeta^{(n-1)})^2.
\end{cases}
$$

This shows that κ_1, κ_2 describes an algebraic curve with a rational parametric representation, see Cox, Little and O'Shea (1992).

Proposition 20. *The uniform distribution on $\{0, \ldots, n-1\}$ has the FGC property, with generating polynomial*

$$
F(\eta, \gamma) = A(B + C)^n - (A + C)B^n,
$$

where $A = \frac{1}{n}\left(\gamma - \eta(1 + \eta)\right)$, $B = \gamma - \eta(n - \eta)$, $C = 1 - n + 2\eta$.

Proof. The moment function is

$$
M(t) = \frac{1}{n}\frac{1 - e^{nt}}{1 - e^t}
$$

and the first two derivatives of the cumulant function are

$$
K'(t) = \frac{-ne^{nt}}{1 - e^{nt}} + \frac{e^t}{1 - e^t},
$$

$$
K''(t) = \frac{-n^2 e^{nt}}{(1 - e^{nt})^2} + \frac{e^t}{(1 - e^t)^2}.
$$

Put

$$
u(t) = \frac{e^{nt}}{1 - e^{nt}}, \quad v(t) = \frac{e^t}{1 - e^t}.
$$

Then

$$
e^{nt} = \frac{u(t)}{1 + u(t)}, \quad e^t = \frac{v(t)}{1 + v(t)},
$$

so

$$
u(t)\left(1 + v(t)\right)^n = (1 + u(t))\, v^n(t) \tag{15}
$$

and

$$
K'(t) = -nu(t) + v(t), \tag{16}
$$

$$
K''(t) = -n^2 u(t)\left(1 + u(t)\right) + v(t)\left(1 + v(t)\right). \tag{17}
$$

Solving (16) and (17) with respect to $K'(t)$, $K''(t)$ by substitution, we get

$$
u(t)(1 - n + 2K'(t)) = \frac{1}{n}\left(K''(t) - K'(t)(1 + K'(t))\right),
$$

$$
v(t)(1 - n + 2K'(t)) = K''(t) - K'(t)(n - K'(t)).
$$

Now substitution in (15) gives the result.

Example 21. Let X have the *mixed exponential distributions*

$$X \sim (1-\alpha)\lambda_1 e^{\lambda_1 x} + \alpha\lambda_2 e^{\lambda_2 x}, \quad 0 < \alpha < 1,$$

for which

$$M_X(\theta) = (1-\alpha)\frac{\lambda_1}{\theta - \lambda_1} + \alpha\frac{\lambda_2}{\theta - \lambda_2}.$$

For $\lambda_1 = 1$ and $\lambda_2 = 2$ and $\alpha = 1/3$ we have the generating polynomial

$$16K'' - 16K'^2 - 44K'^4 - 8K''^2 + K'^6$$
$$+ 44K'^2 K'' - K'^2 K''^2 - K''K'^4 + K''^3.$$

This is easily extended to the following result.

Proposition 22. *Every finite mixture of exponential random variables has the FGC property.*

Proof. As the moment function is a rational function of θ, the derivatives of the cumulant function are rational functions of θ.

The FGC property is preserved under multiplication of the density by a polynomial. We prove the one-dimensional case.

Theorem 23. *Let $p_X(x)$ be the density function of a random variable with the FGC property. Then if Y is a random variable with density $g(y)p_X(y)$ where $g(y)$ is polynomial, Y also has the FGC property.*

Proof. Let $g(y) = \sum_{r=0}^{m} a_r y^r$. Then the moment generating function for Y is $M_Y(t) = \sum_{r=0}^{m} a_r M_X^{(r)}(t)$, where $M_X^{(r)}(t) = \frac{d^r}{dt^r} M_X(t)$. Write $\frac{M_X^{(r)}}{M_X} = U_r$. Then

$$K_Y' = \frac{\sum_{r=0}^{m} a_r U_{r+1}}{a_0 + \sum_{r=1}^{m} U_r}$$

and K_y'' is similarly a rational function involving $U_0, \ldots, U_{r+2}$. The FGC property for X can be written as $F(K_X', K_X'') = 0$ which becomes $F_X(U_1, U_2 - U_1^2) = 0$. Since F is polynomial and $\frac{dU_s}{dt} = U_{s+1} - U_s U_1$, by successive differentiation we can express $U_3, \ldots, U_{m+1}$ in terms of U_1 and U_2. This in turn means that K_Y' and K_Y'' can be expressed in terms of rational functions of U_1 and U_2. These, combined with the original polynomial equation and elimination, lead to a polynomial equation $F_Y(K_Y', K_Y'') = 0$ for Y and hence the FGC property.

The final example in this section is devoted to the discussion of a currently unsolved problem: how to formally check that a given distribution has the FGC property, especially in the case it is known that its cumulant function satisfies

a polynomial differential equation of order strictly greater than 2. In such a case the distribution does not have the FGC property involving only K' and K'', but also needs K''' and possibly higher derivatives. This points to extending the definition to a polynomial equation in the first r derivatives of K and to checking in a constructive way which is the minimal order. We have concentrated on the K', K'' definition here because of its central relationship to the variance function and the fact it is satisfied in most common cases.

Example 24. For the *uniform distribution* on $[0,1]$ the moment generating function is

$$M(\theta) = \frac{e^\theta - 1}{\theta}.$$

Noticing that this involves θ and e^θ, if we set $z = 1/(e^\theta - 1)$ and $t = 1/\theta$, we have the polynomial differential equations $z' = -(1+z)z$, $t' = -t^2$, and we obtain again polynomial differential equations

$$\begin{aligned}
K' &= 1 + z - t, \\
K'' &= -z - z^2 + t^2, \\
K''' &= z + 3z^2 + 2z^3 - 2t^3.
\end{aligned}$$

We eliminate t and z to obtain the generating polynomial

$$\begin{aligned}
(K')^6 &- 5(K')^5 - 3(K')^4 K'' + 17/2(K')^4 + 2(K')^3 K'' - 4(K')^3 K''' \\
&- 6(K')^3 + 3(K')^2(K'')^2 + (K')^2 K'' + 6(K')^2 K''' + 3/2(K')^2 \\
&- 5K'(K'')^2 - 3K'K''' - (K'')^3 + 5/2(K'')^2 - 1/2K'' + 1/2K'''
\end{aligned}$$

Notice that this derivation does not constitute a proof that a lower order equation is not satisfied.

5. The Multivariate Case

The elimination methods of the last section can be used to build systems of random variables related in an algebraic way. As an example consider the the mean function $K'(\theta)$ for the Binomial$(1, 1/2)$ (Bernoulli) and the Poisson(1). Their mean functions are respectively

$$K_1'(\theta) = \frac{e^\theta}{1 + e^\theta}, \quad K_2'(\theta) = e^\theta,$$

giving after elimination of e^θ:

$$K_1'K_2' + K_1' - K_2' = 0,$$

a polynomial relation between K_1' and K_2'.

Definition 25. Two random variables X_1 and X_2 each with their own exponential models and cumulant generating function $K_1(\theta)$ and $K_2(\theta)$ are called *polynomially related* if their differentials lie on a polynomial variety

$$G(K_1'(\theta), K_2'(\theta)) = 0,$$

for all θ in a neighborhood of zero.

This leads to the following transfer of the FGC property from X_1 to X_2.

Theorem 26. *Suppose that X_1 and X_2 are polynomially related and that X_1 has the FGC property. Then so does X_2.*

Proof. Suppose that X_1 and X_2 have cumulant generating functions K_1 and K_2, respectively, satisfying the polynomial variety $G(K_1', K_2')$. Differentiating leads to a polynomial variety in K_1, K_1', K_2, K_2'. We also have one polynomial variety from the FGC properties for X_1: $F_1(K_1', K_1'') = 0$. This gives the right number of equations to eliminate K_1' and K_1'' leaving a single polynomial equation for K_2' and K_2''.

The property of being polynomial related is clearly transitive since if X_1, X_2 and X_2, X_3 are polynomially related pairs we have two polynomial relations for K_1', K_2' and K_2', K_3' and K_2' can be eliminated. Thus the property leads to large equivalence classes of polynomially related random variables. Moreover, by Theorem 26, if one member of such an equivalence class has the FGC property then they all do. The equivalence classes are closed under convolution (see the discussion in the last section) and, by Theorem 23, closed under multiplication of densities by polynomials provided the result is a density.

We can begin to see the outline of large equivalence classes and we give a description of two of them here. We have been unable to completely clarify the totality of each equivalence class but they do not intersect.

Discrete class. This includes all random variables having distributions with finite integer support, but it also includes the Poisson with any mean. It includes all convolutions and polynomially extended distributions. The latter is by the extension of Theorem 23 to discrete distributions. It also includes all distributions on finite rational support.

Normal/Gamma class. Since for the normal $N(\mu, \sigma^2)$ we have $K'(\theta) = \sigma^2\theta + \mu$, a distribution which has K' a rational function of θ will be polynomially related to it. This includes the Gamma(α, λ) for which $K'(\theta) = \alpha\lambda/(\lambda - t)$. Again we can extend by convolution and Theorem 23.

In the case that $X_1, X_2, \ldots$ have a properly defined joint distribution then the above discussion can be interpreted as expressing relations between marginal

distributions. Convolutions are mentioned in Proposition 17, but even though we may not necessarily have independence between X_1 and X_2 we can still make some headway with the following theorem.

Theorem 27. *Let X_1, X_2 be finitely generated random variables with generating polynomials $F_i(\eta_i, \gamma_i)$ and cumulant functions K_i, $i = 1, 2$. Let $\bar{\eta}_i$ denote the mean value and $\bar{\gamma}_i$ the variance of X_i, $i = 1, 2$. For given $a, b \in R$, let $G_{a,b}(\eta_1, \eta_2)$ be a polynomial in $k[\eta_1, \eta_2]$ such that $G_{a,b}(K_1'(at), K_2'(bt)) = 0$ in a neighborhood of 0. Assume that there exists a polynomial $F_{a,b}(\eta, \gamma)$ which is the generator of the elimination ideal*

$$\text{Ideal}\,(F_1(\eta_1, \gamma_1), F_2(\eta_2, \gamma_2), G_{a,b}(\eta_1, \eta_2), f_1, f_2) \cap k[\eta, \gamma],$$
$$f_1 = \eta - a\eta_1 - b\eta_2,$$
$$f_2 = \gamma - a^2\gamma_1 - b^2\gamma_2.$$

If $F_{a,b}$ is a generating polynomial at $(a\bar{\eta}_1 + b\bar{\eta}_2, a^2\bar{\gamma}_1 + b^2\bar{\gamma}_2)$, then $Y = aX_1 + bX_2$ is finitely generated with generating polynomial F.

Remark. All conditions on $F_{a,b}$ are of an algebraic type, and can be checked in a generic way, that is for general a, b, ask if $G_{a,b}$ is a polynomial in the indeterminate a, b, η, γ. Conditions on G are to be checked directly, for example in case of algebraic K_i', $i = 1, 2$, or possibly with differential rings methods on the system of implicit differential equations

$$\begin{cases} F_1(K_1', K_1'') = 0, \\ F_2(K_2', K_2'') = 0. \end{cases}$$

Notice that $G_{a,b}$ is a first integral of a system which is a modification of the previous system, and it will possibly involve $\bar{\eta}_1, \bar{\eta}_2, \bar{\gamma}_1, \bar{\gamma}_2$ in its coefficients as new elements of the basic field k.

Proof. The cumulant function of $Y = aX_1 + X_2$ is given by $K_Y(t) = K_1(at) + K_2(bt)$, so that

$$K_Y'(t) = aK_1'(at) + bK_2'(bt),$$
$$K_Y''(t) = a^2K_1''(at) + b^2K_2''(bt).$$

Let us put $\eta_i = K_i'(at)$, $\gamma_i = K_i''(bt)$, $i = 1, 2$, and $\eta = K_Y'(t)$, $\gamma = K_Y''(t)$. Then the following system of algebraic equations in $k[\eta_1, \eta_2, \gamma_1, \gamma_2, \eta, \gamma]$ is satisfied at least in a neighborhood of $(\bar{\eta}_1, \bar{\eta}_2, \bar{\gamma}_1, \bar{\gamma}_2, a\bar{\eta}_1 + b\bar{\eta}_2, a^2\bar{\gamma}_1 + b^2\bar{\gamma}_2)$:

$$\begin{cases} F_1(\eta - 1, \gamma_1) = 0, \\ F_2(\eta_2, \gamma_2) = 0, \\ G_{a,a}(\eta_1, \eta_2) = 0, \\ \eta - a\eta_1 - b\eta_2\gamma - a^2\gamma_1 - b^2\gamma_2 = 0. \end{cases}$$

The projection of the algebraic curve defined by the five previous equations onto the last two coordinates satisfies the equation $F_{a,b}(\eta, \gamma) = 0$, and in particular $F_{a,b}(K'_Y(t), K''_Y(t)) = 0$ in a suitable neighborhood of 0. As $F_{a,b}$ is a generating polynomial at $(K'_Y(0), K''_Y(0))$, Y is finitely generated by $F_{a,b}$.

Example 28. Let $X_1 \sim \mathrm{Exp}\,(\lambda_1)$, $X_2 \sim \mathrm{Exp}\,(\lambda_2)$. The cumulant functions are

$$K_i(t) = -\ln(1 - \frac{t}{\lambda_i}), \quad i = 1, 2.$$

The first two derivatives are

$$K'_i(t) = \frac{1}{\lambda_i - t}, \quad K''_i(t) = \frac{1}{(\lambda_i - t)^2}, \quad i = 1, 2,$$

and the generating functions are

$$F_1(\eta_1, \gamma_i) = \eta^2 - \gamma, \quad i = 1, 2.$$

For $a, b \in R$ we eliminate t from the system

$$\begin{cases} \eta_1 = \frac{1}{\lambda_1 - at}, \\ \eta_2 = \frac{1}{\lambda_2 - bt}, \end{cases}$$

and obtain

$$G_{a,b}(\eta_1, \eta_2) = (\lambda_1 b - \lambda_2 a)\eta_1\eta_2 + a\eta_1 - b\eta_2.$$

Now we eliminate all unwanted variables from the system

$$\begin{cases} \eta_1{}^2 - \gamma_1 = 0, \\ \eta_2{}^2 - \gamma_2 = 0, \\ (\lambda_1 b - \lambda_2 a)\eta_1\eta_2 + a\eta_1 - b\eta_2 = 0, \\ \eta - a\eta_1 - b\eta_2 = 0, \\ \gamma - a^2\gamma_1 - b^2\gamma_2 = 0. \end{cases}$$

The last polynomial in the Gröbner basis is $F_{a,b}$, that is,

$$\begin{aligned} F_{a,b}(\eta, \gamma) = {}& 4a^2b^2(\eta^2 - 2\gamma) + (\lambda_1 b - \lambda_2 a)^2 \gamma^2 \\ & + (\lambda_1 b - \lambda_2 a)^2 \eta^4 - 2(\lambda_1 b - \lambda_2 a)^2 \eta^2 \gamma. \end{aligned}$$

The partial derivative is

$$\begin{aligned} \frac{\partial}{\partial \gamma} F_{a,b}(\eta, \gamma) &= -8a^2b^2 + 2(\lambda_1 b - \lambda_2 a)^2 \gamma - 2(\lambda_1 b - \lambda_2 a)^2 \eta^2 \\ &= -8a^2b^2 + 2(\lambda_1 b - \lambda_2 a)^2(\gamma - \eta^2). \end{aligned}$$

The multivariate definition of the FGC property is not completely satisfactory because we do not know if it implies the FGC property for all linear

transformations. The following definition is intended to clarify a more restrictive property.

Definition 29. The random vector $X = (X_1, X_2)$ is jointly finitely generated if for all $a, b \in R$ such that $a^2 + b^2 = 1$ the random variable $aX_1 + bX_2$ is finitely generated.

Proposition 30. *Let the random vector X be jointly finitely generated and let $a, b \in R$ with $\rho = \sqrt{a^2 + b^2} \neq 0$. Then the random variable $aX_1 + bX_2$ is finitely generated with generating polynomial $F(\rho^{-1}\eta, \rho^{-2}\gamma)$, where $F(\eta, \gamma)$ is the generating polynomial of $\frac{a}{\rho}X_1 + \frac{b}{\rho}X_2$.*

Proof. If $Y = \frac{a}{\rho}X_1 + \frac{b}{\rho}X_2$ and K_Y is the cumulant function of Y, then $K_{\rho Y}(t) = K_Y(\rho t)$ and $K'_{\rho Y}(t) = \rho K'_Y(\rho t)$, $K''_{\rho Y}(t) = \rho^2 K_Y(\rho t)$. Then

$$F(\rho^{-1}K'_{\rho Y}(t), \rho^{-2}K''_{\rho Y}(t)) = F(K'_Y(\rho t), K''_Y(\rho t)) = 0.$$

Finally the polynomial $F(\rho^{-1}\eta, \rho^{-2}\gamma)$ is a generating polynomial for all $\rho \neq 0$.

6. Applications

6.1. Generalised linear models

In the statistical use of the exponential family we assume that Y comes from an exponential family

$$Y \sim p(y, \theta) = e^{\theta y - K_Y(\theta)} p_0(y),$$

where $Y_0 \sim p_0(y)$. Then θ is called the natural parameter and is typically used in modeling. For example when there are covariates z we model θ as a function of z and the model parameters β, $\theta = f(z, \beta)$. In this framework the mean and variance of Y_θ are given respectively by $\eta_\theta = K'(\theta)$, $\sigma_\theta^2 = K''(\theta)$, and the variance function has its usual interpretation $\sigma_\theta^2 = V(\eta_\theta)$. This means that the implicit representation in the FGC property can be interpreted as $F(\eta_\theta, \sigma_\theta^2) = 0$ and in the modeling case we have $F(K'(f(z, \beta)), K''(f(z, \beta))) = 0$.

In the sample case we observe $Y_{\theta_1}, \ldots, Y_{\theta_n}$ where $\theta_i = f(z_i, \beta)$. Assume also the linear case, namely $\theta_i = f(z_i, \beta) = \sum_{j=1}^m p_j(z_i)\beta_j$ for functions $p_j(z)$, for example monomials. In this case the log-likelihood is, up to a constant

$$l = \sum_{i=1}^n (\theta_i y_i - K(\theta_i))$$

and the likelihood equations are

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n (y_i - K'(\theta_i)) \, p_j(z_i).$$

The usual information matrix is

$$\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n} p_j(z_i) p_k(z_i) K''(\theta_i).$$

These equations are then combined with $F(K'(\theta_i), K''(\theta_i)) = 0$ to obtain a full set of equations for much standard statistical inference. This has computational advantages over explicitly computing $\hat{\theta}$ and substituting to obtain $K^{(n)}(\hat{\theta})$.

Returning to the one-sample case, since development of $K^{(r)}(\theta)$ is possible by successive differentiation of F, we can similarly develop the empirical MLE-version by differentiation and substituting $y = K'$.

6.2. Saddle point approximations

An application of this is to saddle-point approximation. The approximation is based on the exponential model as described here. The random variable X of interest has its density embedded in the exponential model as described above. For an i.i.d. sample $X_1, \ldots, X_n$ we define the partial sums $S_n = X_1 + \cdots + X_n$. This has density

$$p_{S_n}(s; \theta) = e^{\theta s - n K_X(\theta)} p_{S_n}(s)$$

for any θ. The saddle-point approximation to p_{S_n} is given by

$$p_{S_n} = \frac{e^{-\hat{\theta}s + n K_X(\hat{\theta})}}{2\pi n K_X''(\hat{\theta})^{\frac{1}{2}}} \{1 + O(n^{-1})\},$$

where $\hat{\theta}$ is the maximum likelihood estimator given by the solution of $K'(\hat{\theta}) = s/n$. Again the approach taken here can give precise relationships for K'' by replacing K' by s/n in the formula $F(K', K'') = 0$. The term $nK''(\hat{\theta})$ is sometimes referred to as the observed information. This can be extended to the modeling case in a straightforward manner.

6.3. Edgeworth expansions

Theorem 23 has application to Edgeworth expansions. A natural place where we have densities or approximations to densities is in a formal Edgeworth expansion:

$$\psi(x) = \sum_{r=0}^{n} a_r H_r(x) \phi(x),$$

where $H_r(x)$ is the rth Hermite polynomial and $\phi(x)$ is the standard Normal density. Suppose $\psi(x)$ is an actual density and define a polynomial $g(\theta) =$

$\sum_{r=0}^{n} a_r \theta^r$. The moment generating function for the density $\psi(x)$ is

$$M(t) = \sum_{r=0}^{n} \int_{-\infty}^{+\infty} H_r(x) e^{tx} \phi(x) dx$$
$$= \sum_{r=0}^{n} a_r t^r e^{\frac{1}{2}t^2} = g(t) e^{\frac{1}{2}t^2}.$$

Then for the derivatives of the cumulant function

$$K(t) = \log g(t) + \frac{1}{2}t^2,$$
$$K'(t) = \frac{g'(t)}{g(t)} + t,$$
$$K''(t) = \frac{g''(t)}{(g(t))^2} - \left(\frac{g'(t)}{g(t)}\right)^2 + 1.$$

So, elimination will lead to a polynomial in K', K'', as is expected from Theorem 23.

References

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* John Wiley, New York.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman & Hall, London.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics.* Chapman & Hall, London.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Lecture Notes - Monograph Series.* Vol. 9. Institute of Mathematical Statistics, Hayward, California.

Capani, A., Niesi, G. and Robbiano, L. (1995). CoCoA, a system for doing computations in commutative algebra. Available via anonymous ftp from `lancelot.dima.unige.it`.

Casalis, M. (1996) The $2d+4$ simple quadratic natural exponential families on **R**. *Ann. Statist.* **24**, 1828-1854.

Char, B., Geddes, K., Gonnet, G., Leong, B., Monogan, M. and Watt, S. (1991). *MAPLE V Library Reference Manual.* Springer-Verlag, New York.

Cox, D., Little, J. and O'Shea, D. (1992) *Ideal, Varieties, and Algorithms.* Springer-Verlag, New York.

Feller, W. (1966) *An Introduction to Probability Theory and Its Applications.* Vol. 2. John Wiley, New York.

Gibilisco, P. and Pistone, G. (1998). Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **1**, 325-347.

Gutiérrez-Peña, E. and Smith, A. F. M. (1997). Exponential and Bayesian conjugate priors: Review and extensions. *Test* **6**, 1-90.

Johansen, S. (1979). *Introduction to the Theory of Regular Exponential Families. Lecture Notes.* Vol. 3. Institute of Mathematical Statistics, University of Copenhagen.

Kendall, W. S. (1993). Computer algebra in probability and statistics. *Statistica Neerlandica* **47**, 2-25.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. Dover, New York.

Letac, G. (1992). *Lectures on Natural Exponential Families and Their Variance Functions*. Number 50 in "Monografias de Matemática", Istituto de Matemática Pura e Aplicada, Rio de Janeiro.

Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Ann. Statist.* **18**, 1-37.

McCullagh, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika* **71**, 461-476.

McCullagh, P. (1987). *Tensor Methods in Statistics*. Monographs in Statistics and Applied Probability. Chapman & Hall, London.

Mora, M. (1986). Classification de fonctions variance cubiques des familles exponentielles sur $\boldsymbol{R}^n$. *Comptes Rendus de l'Academie de Sciences de Paris Série I Mathematiques* **302**, 587-590.

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

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.* **11**, 515-529.

Pistone, G. and Rogantin, M.-P. (1998). The exponential statistical manifold: mean parameters, orthogonality, and space transformation. *Bernoulli*. To appear.

Pistone, G. and Sempi, C. (1995) An infinite dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.* **33**, 1543-1561.

Pistone, G. and Wynn, H. P. (1996). Generalised confounding with Gröbner bases. *Biometrika* **83**, 653-666.

Department of Mathematics, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy.

E-mail: pistone@calvino.polito.it

Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom.

E-mail: hpw@stats.warwick.ac.uk