SAMSI

2008-09 Program on Algebraic Methods in Systems Biology and Statistics

Algebraic features of cumulants

Giovanni Pistone Henry Wynn

SAMSI RTP NC

LSE London UK

Feb 2nd, 2009

Summary

Generalities Multivariate cumulants: definition, existence, combinatorial properties, log-exp formulæ. Exponential models, divergences.

Finite state space Monomial aliasing, moments aliasing, cumulants aliasing, polynomial representation. Models on the finite state space.

Finite generation Differential finite generation of cumulants.

Geometric approach Interpretation of cumulants on the statistical manifold.

Multivariate cumulants I

Definition (Moment and cumulant generating function)

- X is a random vector in $\mathbb{R}^m$.
- For $\theta \in \mathbb{R}^m$, $\theta \cdot X = \sum_{i=1}^m \theta_i X_i$ is the scalar product.
- D_X is the *interior* of the convex set

$$\left\{\theta \in \mathbb{R}^m : E[e^{\theta \cdot X}] < +\infty\right\}.$$

- 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 the equations

$$M_X(\theta) = E[e^{\theta \cdot X}],$$
$$K_X(\theta) = \log M_X(\theta).$$

If $D_X = \emptyset$, one could consider the characteristic function $\Phi_X(\theta)$ and define K_X by $\mathrm{e}^{K_X(\theta)} = \Phi_X(\theta)$, $K_X(0) = 0$. Cf. Eugene Lukacs. *Characteristic functions*. Hafner Publishing Co., New York, 1970. Second edition, revised and enlarged.

Multivariate cumulants II

Theorem (Convexity and analyticity)

1. *The moment generating function* M_X *and the cumulant generating function* K_X *are convex.*
2. *If* X *is not supported by an affine hyperplane of dimension* $< m$*, then the convexity is strict.*
3. *The moment function* M_X *and the cumulant function* K_X *are analytic in* D_X*. The* α*-derivatives of the moment generating function are given by*

$$D^\alpha M_X(\theta) = \frac{\partial^{\alpha_1+\cdots+\alpha_m}}{\partial\theta_1^{\alpha_1}\ldots\partial\theta_m^{\alpha_m}} M_X(\theta) = E[X_1^{\alpha_1}\cdots X_m^{\alpha_m} e^{\theta\cdot X}] = E[X^\alpha e^{\theta\cdot X}].$$

4. *If the generating functions are defined in 0, then the random vector* X *has finite moments of all orders given by*

$$\mu_\alpha(X) = D^\alpha M_X(\theta)|_{t=0} \quad .$$

Multivariate cumulants III

Theorem (Monotonicity)

- *The gradient of the cumulant generating function*

$$\nabla K_X(\theta) = (\frac{\partial}{\partial \theta_1} K_X(\theta), \ldots, \frac{\partial}{\partial \theta_m} K_X(\theta))$$

is a (stritly) monotone function D_X to $\mathbb{R}^m$, i.e.

$$\langle \nabla K_X(\theta_1) - \nabla K_X(\theta_2), \theta_1 - \theta_2 \rangle > 0, \quad \theta_1 \neq \theta_2$$

- *In particular, ∇K_X is a 1-to-1 analytic mapping from D_X to $\mathbb{R}^m$.*
- *For each matrix $A \in \mathbb{R}^{m \times m}$ the initial value problem*

$$\dot{\theta}(t) = -\nabla K_X(\theta(t)) + A\theta(t)$$

has a unique solutions such that

$$\|\theta_1(t) - \theta_2(t)\|^2 \leq \|\theta_1(0) - \theta_2(0)\|^2 \, \mathrm{e}^{\|A\| t}$$

Multivariate cumulants IV

Definition (Cumulants)

1. 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 L^{e} of exponentially integrable random vectors.
2. If $X \in L^{\mathrm{e}}$, then all joint moments are defined:

$$\mu_\alpha = \mathsf{E}\left(X^\alpha\right), \quad \mu'_\alpha = \mathsf{E}\left(\left(X - \mathsf{E}\left(X\right)\right)^\alpha\right)$$

3. If $X \in L^e$, the α-derivatives of K are called cumulants:

$$\kappa_\alpha(X) = D^\alpha K_X(\theta)|_{\theta=0}$$

and the Taylor expansion of K_X is

$$K_X(\theta) = \sum_\alpha \frac{\kappa_\alpha(X)}{\alpha!}\theta^\alpha.$$

Applications

Sample For n independent random vectors $X_1, \ldots, X_n$, $X = (X_1, \ldots, X_N)$ in $\mathbb{R}^{\sum m_j}$, the joint CGF is the *outer* sum:

$$K_X(\sigma_1, \ldots, \sigma_n) = K_{X_1}(\sigma_1) + \cdots + K_{X_n}(\sigma_n),$$

Sum For n independent random vectors $X_1, \ldots, X_n$ the joint CGF of the (inner) sum $Y = \sum_{j=1}^n X_j$ is the sum of the marginal CGFs:

$$K_Y(\theta) = K_{X_1}(\theta) + \cdots + K_{X_n}(\theta)$$

$\otimes$ Independence of $X_1, \ldots, X_n$ is equivalent to all mixed cumulants being 0.

ICA/CCA For n independent random variable $X_1, \ldots, X_n$ and a matrix $A \in \mathbb{R}^{m,n}$, with columns $a_1, \ldots, a_m$, the joint CGF of the random vector $Y = AX$ is

$$K_Y(\theta) = K_X(A^{\text{transpose}}\theta) = K_{X_1}(a_1 \cdot \theta) + \cdots + K_{X_m}(a_m \cdot \theta)$$

- Lek-Heng Lim and Jason Morton. Cumulant component analysis: a symultaneous generalization of PCA and ICA. Computational Algebraic Statistics, Theories and Applications, December 10-11, 2008, Kyoto, 2008

Existence

The use of generating functions, moments, cumulants raises non trivial question of existence and uniqueness. We want:

- Characterise the class of moment generating functions.
- Characterize the class of cumulant generating functions.
- Characterise sequences that are moments of a unique distribution.
- Characterise sequence that are cumulants of a unique distribution.
- Characterize cumulants of a given order.
  2. Let X be an m-dimensional random vector. If α is a multi-exponent of degree 2, e.g. $110, 101, 011, 200, 020, 002$, then $\kappa_\alpha(X)$ is a symmetric 2-tensor that we identify with the covariance matrix: $\kappa_\alpha(X) = \operatorname{Cov}(X_i, X_j)$, $i, j : \alpha_i \alpha_j \neq 0$. Viceversa, given a symmetric 2-tensor $K_{i,j}$, $i, j = 1 \dots m$, it is the cumulant of some distribution if and only if it is non-negative definite.
  3. If α is a multi-exponent of degree 3, we have the symmetric 3-tensor $K_{ijk} = \kappa^\alpha$, $\alpha_i \alpha_j \alpha_k \neq 0$. Viceversa ...
- Is the class of ICA distributions general?

Moments, cumulants, κ-statistics

1. From $M_X(\theta) = \mathrm{e}^{K_X(\theta)}$ and the Faà di Bruno formula one shows that the moment μ_α is a symmetric polinomial of the cumulants $\kappa_\beta(X)$ with $\beta \leq \alpha$.
2. From $K_X(\theta) = \log\left(K_X(\theta)\right)$ and the Faà di Bruno formula one shows that the cumulant $\kappa_\alpha(X)$ is a symmetric polinomial of the moments $\mu_\beta(X)$ with $\beta \leq \alpha$.
3. The k-statistics of oder α is the (unique) symmetric polynomial of an independent sample of X such that the expected value is $\kappa_\alpha(X)$.

- Peter McCullagh. *Tensor Methods in Statistics*. Monographs in Statistics and Applied Probability. Chapman & Hall, London, 1987[Ch. 4]
- Francesco Vaccarino. The ring of multisymmetric functions. *Ann. Inst. Fourier (Grenoble)*, 55(3):717–731, 2005. ISSN 0373-0956
- Michael Hardy. Combinatorics of partial derivatives. *Electron. J. Combin.*, 13(1):Research Paper 1, 13 pp. (electronic), 2006. ISSN 1077-8926
- Elvira Di Nardo, Giuseppe Guarino, and Domenico Senato. A unifying framework for k-statistics, polykays and their multivariate generalizations. *Bernoulli*, 14(2):440–468, 2008

Binary moments and cumulants

- A binary exponent is a coding of a subset of indices, e.g. 11100 is the subset $1 \cdots 3$ of the set $1 \cdots 5$. A *partition* b of α is a set $\{\alpha_1, \ldots, \alpha_k\}$ such that $\sum_j \alpha_j = \alpha$. E.g., for $n = 3$ we have 1|2|3, 12|3, 1|23, 13|2, 123, or $11100 = 10000 + 01000 + 00100$, $11100 = 11000 + 00100 \ldots$. These partitions together with the refinement partial order $\prec$ define a lattice: e.g. $12|3|45 \prec 12|345$.
- Each such binary code has a corresponding moment, centered moment and cumulant, e.g. μ_{111000}, μ'_{111000}, κ_{111000}. The product over all terms in a partition is is denoted $\prod_{\alpha \in b}$.
- For a binary α, the Faà di Bruno formulæ, applied to exp and log give:

$$\mu_\alpha = \sum_{b \prec \alpha} \prod_{\beta \in b} \kappa_\beta$$

$$\kappa_\alpha = \sum_{b \prec \alpha} (-1)^{|b|-1} (|b| - 1)! \prod_{\beta \in b} \mu_\beta$$

where $|b|$ is the number of elements of the partition b.

Example of log-exp formulæ

As $\mu'_\alpha(X) = \sum_{\beta \leq \alpha} \mu_\beta(X) \prod_{i:\ \alpha_i=1, \beta_i=0} (-\mu_i(X))$, we have:

- $$\begin{aligned} \mu'_{111} = \mu_{111} - (\mu_{110}\mu_{001} + \mu_{101}\mu_{010} + \mu_{011}\mu_{100}) + \\ (\mu_{100}\mu_{010}\mu_{001} + \mu_{010}\mu_{100}\mu_{001} + \mu_{001}\mu_{100}\mu_{010}) + \mu_{100}\mu_{010}\mu_{001} = \\ \mu_{111} - (\mu_{110}\mu_{001} + \mu_{101}\mu_{010} + \mu_{011}\mu_{100}) + 2\mu_{100}\mu_{010}\mu_{001} \end{aligned}$$
- $$\kappa_{111} = \mu_{111} - \mu_{110}\mu_{001} - \mu_{101}\mu_{010} - \mu_{011}\mu_{100} + 2\mu_{100}\mu_{010}\mu_{001}$$
- Therefore, $\mu'_{111} = \kappa_{111}$.

- Giovanni Pistone and Henry P. Wynn. Cumulant varieties. *Journal of Symbolic Computation*, 41(2):210–221, 2006. ISSN 0747-7171
- Ole E. Barndorff-Nielsen and D. R. Cox. *Asymptotic techniques for use in statistics*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1989. ISBN 0-412-31400-2.

Monomial and moment aliasing

- Let D be a finite set of points in $\mathbb{R}^m$, $I(D)$ the design ideal in $\mathbb{R}[x_1, \ldots, x_m]$, $g_1, \ldots, g_k$ a polinomial basis of $I(D)$, x^α, $\alpha \in L$, a linear monomial basis of $\mathbb{R}[x_1, \ldots, x_m]/I(D)$. This is the usual setting of the algebric theory of Design of Experiments. Each equation $g(x) = 0$, $g \in I(D)$, is an *aliasing* relation between moments.
- Let
$$H(x) = \exp\left(\sum_{i=1}^{n} s_i x_i\right) = \sum_{\alpha \in L} b_\alpha(s) x^\alpha.$$
Therefore $M_X(s) = \sum_{\beta \geq 0} \frac{s^\beta \mu_\beta}{\beta!} = \sum_{\alpha \in L} b_\alpha(s) \mu_\alpha$, so that
$$\mu_\beta = \sum_{\alpha \in L} b_{\alpha,\beta} \mu_\alpha,$$
where $b_{\alpha,\beta} = D_\beta b_\alpha(s)|_{s=0}$.
- Note that these coefficients, coming as they do from the initial interpolation, only depend on the support and choice of the monomial basis.

Cumulant aliasing

For a discrete distribution and monomial order τ every cumulant $\mu_\beta, \beta \geq 0$ is expressible as a linear function of the moments $\mu_\alpha,\ \alpha \in L$, whose coefficients depend only the support and choice of monomial ordering, not the $p(x)$.

Theorem (Cumulants aliasing)

For a discrete distribution and monomial order τ every cumulate $\kappa_\beta, \beta \geq 0$ is expressible as a polynomial function of the cumulates $\kappa_\alpha,\ \alpha \in L$, whose form is only dependent of the support and monomial ordering.

Proof. Express κ_β as polynomial in moments $\mu_\beta \leq \alpha$. Then use the moment aliasing to express any such μ_β in terms of the $\mu_\gamma, \gamma \in L$. Then express any such μ_γ in terms of $\kappa_\delta, \delta \leq \gamma$. The order ideal property of L means that all such δ remain in L.

- Giovanni Pistone and Henry P. Wynn. Cumulant varieties. *Journal of Symbolic Computation*, 41(2):210–221, 2006. ISSN 0747-7171
- Giovanni Pistone, Eva Riccomagno, and Henry P. Wynn. *Algebraic statistics*, volume 89 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL, 2001. ISBN 1-58488-204-2. Computational commutative algebra in statistics

Example of cumulant aliasing

Let $m = 2$ and take $D = \{0,1,2\}^2 \setminus \{(2,2)\}$, i.e.

$$\{(0,0),(1,0),(0,1),(2,0),(1,1),(0,2),(2,1),(1,2)\}.$$

From the special structure of D we can claim that $D = L$ for any monomial ordering. Note that $(2,2)$ is not in L and our aim is the express κ_{22} in terms of $\kappa_\alpha,\ \alpha \in L$.
First express κ_{22} in terms of moments:

$$\begin{aligned} k_{22} \quad = \quad & \mu_{22} - 2\mu_{21}\mu_{01} + 2\mu_{20}\mu_{01}^2 \\ & -\mu_{20}\mu_{02} - 2\mu_{11}^2 + 8\mu_{10}\mu_{11}\mu_{01} \\ & -2\mu_{10}\mu_{12} - 6\mu_{10}^2\mu_{01}^2 \\ & +2\mu_{10}^2\mu_{02} \end{aligned}$$

But we can fold m_{22} in to L. One method for doing this in simple problems is to use the Gröboner basis directly.

The relevant G-basis element is

$$x_1^2 x_2^2 - x_1^2 x_2 - x_1 x_2^2 + x_1 x_2$$

Since this is zero on the support D we can take expectation directly to obtain

$$m_{22} = m_{21} + m_{12} - m_{11}$$

Substituting for m_{22} and converting back to cumulates we have the following value for κ_{22} :

$$\begin{gathered}
-\kappa_{11} + \kappa_{21} + \kappa_{12} + \kappa_{10}\kappa_{01} - \kappa_{20}\kappa_{02} + \kappa_{10}^2\kappa_{01} \\
+\kappa_{10}\kappa_{01}^2 + \kappa_{20}\kappa_{01} - \kappa_{10}\kappa_{02} - \kappa_{10}^2\kappa_{02} \\
-\kappa_{20}\kappa_{01}^2 + 2\kappa_{10}\kappa_{11} + 2\kappa_{01}\kappa_{11} - 2\kappa_{10}\kappa_{12} \\
+2\kappa_{10}\kappa_{21} - 2\kappa_{11}^2 + 4\kappa_{10}\kappa_{01}\kappa_{11} - 17\kappa_{10}^2\kappa_{01}^2
\end{gathered}$$

Exponential models

Definition (Exponential model)

1. Let X be a random vector of class L^{e} and let D_X be the domain of the generating functions. Then the equation

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

defines a statistical model.

2. Such a model is called the *natural exponential model associated to* X.
3. The model can be parameterized by the *mean parameter* $\eta = K_X'(\theta)$

Finite generation I

Definition

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

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

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

$$\eta_0 = K_X'(0) \quad , \quad \gamma_0 = K_X''(0) \quad ,$$

and the equations

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

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

Finite generation II

Definition (Variance function)

Let Ψ_X be the inverse of the gradient of the cumulant generating function K_X. The *variance function* of the exponential model is

$$V_X(\eta) = K_X''[\Psi_X(\eta)] \quad ,$$

where K_X'' is the Hessian matrix of the cumulant generating function.

A number of authors have classified distributions for which the variance function $V_X(\cdot)$ has an algebraic form.

- Carl N. Morris. Natural exponential families with quadratic variance functions. *The Annals of Statistics*, 10:65–80, 1982
- Carl N. Morris. Natural exponential families with quadratic variance functions: Statistical theory. *The Annals of Statistics*, 11:515–529, 1983
- Marianne Mora. Classification de fonctions variance cubiques des familles exponentielles sur $\mathbb{R}^n$. *Comptes Rendus de l'Academie de Sciences de Paris Série I Mathematiques*, 302:587–590, 1986
- Gérard Letac and Marianne Mora. Natural real exponential families with cubic variance functions. *The Annals of Statistics*, 18:1–37, 1990

Finite generation III

- We refer to the property in the Definition 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 $(K_X''(0), K_X'(0))$.
- 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 give a algorithm for generation of the coefficients of the Taylor development of $K(t)$ at t_0.
- 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.

- Giovanni Pistone and Henry P. Wynn. Finitely generated cumulants. *Statist. Sinica*, 9(4):1029–1052, October 1999. ISSN 1017-0405

Variance function: Morris

- The following table is adapted from [Morris, 1982, Table 1], where all the distributions such as the variance function is a quadratic polynomial in the mean are studied.
- In our terms, the variance $K''(\theta)$ and the mean $K'(\theta)$ are related by a *generating polynomial* of degree 2.

Distribution	Parameters	Generating polynomial
Normal $\mathrm{N}(\mu, \sigma^2)$	$\mu,\ \sigma^2$	$K'' - \sigma^2$
Poisson $\mathrm{P}(\lambda)$	λ	$K'' - K'$
Gamma $\Gamma(\alpha, \lambda)$	$\alpha,\ \lambda$	$\alpha K'' - (K')^2$
Binomial $\mathrm{Bin}(n, p)$	$n,\ p$	$nK'' - K'(n - K')$
Negative Binomial $\mathrm{NegBin}(r, p)$	$r,\ p$	$rK'' - K'(r + K')$
Generalised Hyperbolic Secant	$r,\ \lambda = \tan t$	$rK'' - (K')^2 - r^2$

Finite generation IV

- The generating polynomial uniquely defines the corresponding distribution. E.g. the differential equation for $\eta(\theta) = K'(\theta)$ in the GHS case is

$$r\eta'(\theta) = \eta(\theta)^2 + r^2, \quad \eta(0) = 0$$

  The unique solution is

$$\eta(\theta) = r \tan t$$

  so that

$$K(\theta) = r \int_0^t \tan \tau \, d\tau = r \log \sec t.$$

- All cumulants are polynomials in the mean parameter. E.g. for the GHS distribution

$$r^n K^{(n)}(\theta) = f_n(K'(\theta)), n = 2, 3, \ldots$$

  where

$$f_{n+1}(\eta) = f_n'(\eta)(\eta^2 + r^2).$$

Finite generation V

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

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

The generating polynomial is

$$(K'')^2 - 2(1 + (K')^2)K'' + (K')^2 + (K')^4.$$

- The uniform density on $\{0, 1, 2\}$ has generating polynomial

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

Finite generation: properties

Theorem

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.*

Theorem

- *Every discrete distribution supported on an equally spaced set of reals has the FGC property.*
- *Every finite mixture of exponential random variables has the FGC property.*
- *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 then Y also has the FGC property.*

Finite generation: discussion

- For U[0, 1] the MGF is $M(\theta) = \frac{e^\theta - 1}{\theta}$.
- This involves θ and e^θ. We set $z = \frac{1}{e^\theta - 1}$ and $t = \frac{1}{\theta}$, so that $z' = -(1+z)z \;\; t' = -t^2$ and

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

- Algebraic elimination of t and z gives

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

- This derivation does not constitute a proof that a lower order equation is not satisfied. Differential algebra?

Statistical manifold

- Let $\mathcal{M}$ be the set of *all* probability densities on a probability space. At each density $p \in \mathcal{M}$ we attach a *tangent space* T_p consisting of p-centered random variables, i.e. $u \in T_p$ if $\mathrm{E}_p[u] = \int up\, d\mu = 0$.
- Each other density $q \in \mathcal{M}$ is represented, with respect to p, by a unique element u of the tangent space T_p such that

$$q = \mathrm{e}^{u - K_p(u)} \cdot p$$

$$u = \log\left(\frac{q}{p}\right) - \mathrm{E}_p\left[\log\left(\frac{q}{p}\right)\right]$$

- The mapping $q \to u$ is the p-chart in the statistical manifold and we denote the expectation w.r.t. q by $\mathrm{E}_u[v] = \int vq\, d\mu$.

- Giovanni Pistone and Carlo Sempi. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Statist.*, 23(5):1543–1561, October 1995. ISSN 0090-5364
- Paolo Gibilisco and Giovanni Pistone. Connections on non-parametric statistical manifolds by Orlicz space geometry. *IDAQP*, 1(2):325–347, 1998. ISSN 0219-0257
- Giovanni Pistone and Maria Piera Rogantin. The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli*, 5(4):721–760, August 1999. ISSN 1350-7265
- Alberto Cena and Giovanni Pistone. Exponential statistical manifold. *AISM*, 59:27–56, 2007. ISSN 0020-3157. doi10.1007/s10463-006-0096-y. On line since December 16, 2006
- Giovanni Pistone. Algebraic varieties vs differentiable manifolds in statistical models. Chapter in a forthcoming book edited by P. Gibilisco, E. Riccomagno, M.-P. Rogantin, H. Wynn, 2008

Moment functional and cumulant functional

Theorem

1. *From* $q = \mathrm{e}^{u - K_p(u)} \cdot p = \frac{\mathrm{e}^u}{\mathrm{e}^{K_p(u)}} \cdot p$ *and the condition* $\int q \, d\mu = 1$*:*

$$M_p(u) = \mathrm{E}_p\left[\mathrm{e}^u\right]$$
$$K_p(u) = \log\left(M_p(u)\right)$$

2. $D(p\|q) = \mathrm{E}_p\left[\log\left(\frac{p}{q}\right)\right] = \mathrm{E}_p\left[K_p(u) - u\right] = K_p(u).$
3. M_p *and* K_p *are both convex functionals on* T_p*, with values in* $\mathbb{R}_{>0} \cup \{+\infty\}$.
4. M_p *and* K_p *are analytic functionals on the interior of the set* $\{u \in T_p : M_p(u) < +\infty\}$. *It is a convex open domain* $\mathcal{S}_p \subset T_p$.
5. $\mathcal{E}_p = \left\{q \in \mathcal{M} : q = \mathrm{e}^{u - K_p(u)} \cdot p, u \in \mathcal{S}_p\right\}$ *is the maximal exponential model at* p.

Directional derivatives

- For each analytic functional $f : \mathcal{S}_p \to \mathbb{R}$ the n-th derivative at u, $D^n f(u)$ is a *symmetric multilinear operator* (tensor) such that for each $v_1, \dots, v_n \in T_p$,

$$D^n f(u)(v_1, \dots, v_n) = \frac{\partial^n}{\partial t_1 \cdots \partial t_n} f(u + t_1 v_1 + \cdots + t_n v_n)\bigg|_{t=0}$$

- As $D^n f(u)$ is symmetric, $(v_1, \dots, v_n)$ can be treated as *commutative indeterminates* and written $(v_1 \circ \dots \circ v_n)$. Symmetric n-operators can be mappend into symmetric $(n+1)$-operators.
- $\mu_n(u) v_1 \circ \dots \circ v_n = D^n M_p(u)(v_1 \circ \dots \circ v_n) = \mathrm{E}_p\left[v_1 \cdots v_n \mathrm{e}^u\right]$.
- $\kappa_n(u) v_1 \circ \dots \circ v_n = D^n K_p(u)(v_1 \circ \dots \circ v_n)$.

- J. Dieudonné. *Foundations of Modern Analysis*. Academic press, New York, 1960
- Serge Lang. *Differential and Riemannian manifolds*, volume 160 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 1995. ISBN 0-387-94338-2

Recursion

- Derivatives are computed recursively,

$$D^n f(u) v_1 \circ \cdots v_n = \frac{d}{dt} D^{n-1} f(u + t v_n) v_1 \circ \cdots v_{n-1} \Big|_{t=0}$$

- $DK_p(u)v = \frac{d}{dt} \log\left(\mathrm{E}_p\left[\mathrm{e}^{u+tv}\right]\right)\Big|_{t=0} = \mathrm{E}_u\left[v\right]$
-

$$D^2 K_p(u) v_1 \circ v_2 = \frac{d}{dt} \mathrm{E}_{u+tv_2}\left[v_1\right]\Big|_{t=0} = \frac{d}{dt} \mathrm{E}_p\left[v_1 \mathrm{e}^{u+tv_2-K_p(u+tv_2)}\right]\Big|_{t=0} = \mathrm{E}_u\left[v_1(v_2 - \mathrm{E}_u\left[v_2\right])\right].$$

-

$$\kappa_n(u) v_1 \circ \cdots v_n = D^n K_p(u) v_1 \circ \cdots v_n = \frac{d}{dt} \kappa_{n-1}(u + t v_n) v_1 \circ \cdots v_{n-1} \Big|_{t=0}$$

Exponential models

- An exponential model $\mathcal{E}_V$, where V is a linear subspace on T_p is the model $\mathrm{e}^{u-K_p(u)} \cdot p$, $u \in V$.
- The set of finite-dimensional exponential models is the set of finite-dimensional linear subspaces of the tangent space T_p.
- The restriction of the system of cumulants $\kappa_n(u)v_1 \circ \cdots \circ v_n$, $n = 1, 2, \dots,$ $u, v_1, \dots, v_n \in V$, is computed locally on the exponential model $\mathcal{E}_V$.
- Higher order cumulants are computed as $\kappa_{|\alpha|}(u)(v_1^{\circ\alpha_1} \circ \cdots \circ v_n^{\circ\alpha_n}$, $u, v_1, \dots, v_n \in V$.