7th World Congress in Probability and Statistics
Singapore, July 14 - 19, 2008
Jointly sponsored by the Bernoulli Society and the Institute of Mathematical Statistics

Umbral methods in statistics

Elvira Di Nardo
(Univ. Basilicata – Italy)
www.unibas.it/utenti/dinardo/home.html

Tensor Methods in Statistics

(P. Mc Cullagh) 1987

- *In matters of aesthetics and mathematical notation, no one loves an index. According to one school of thought, indices are the pawns of an arcane and archaic notation, the front-line troops, the cannon fodder, first to perish in the confrontation of an inner product. Only their shadows persist.*

ONLY THEIR SHADOWS PERSIST

Umbrae

$$(\alpha + \gamma)^n \longrightarrow \sum_{i=0}^{n} \binom{n}{i} a_i \, g_{n-i}$$

$$\begin{cases} \alpha \text{ "represents" } \{a_i\} \\ \gamma \text{ "represents" } \{g_i\} \end{cases}$$

$$\sum_{k_1+k_2+\cdots+k_m=n} \binom{n}{k_1, k_2, \ldots, k_m} a_{k_1} a_{k_2} \cdots a_{k_m} \longleftarrow (m \cdot \alpha)^n$$

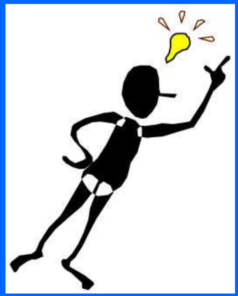
***n*-th moment of a sum of *m* i.i.d. random variables**

Tensor Methods in Statistics
(P. Mc Cullagh) 1987

- *This extreme scorn and derision for indices is not universal. It can be argue, for example, that a "plain unadorned letter" conceals more than it reveals. Like the grin on Alice's cat, the indices can remain long after the symbol has gone. Just as the grin rather than the Cat is the visible display of the Cat's disposition, so too it is the index, not the symbol, that is the visible display of the nature of the mathematical object.*

α "represents" $\{a_i\}$

$\alpha^i \xrightarrow{?} a_i$



How?

$$\alpha^i \, \gamma^{n-i}$$

$$\sum_{i=0}^{n} \binom{n}{i} a_i \, g_{n-i} \quad \Rightarrow \quad (\alpha+\gamma)^n$$

the main idea

Example:

$$\sum_{n=0}^{\infty} B_n \frac{x^n}{n!} = \frac{x}{e^x - 1}$$

B_n Bernoulli numbers

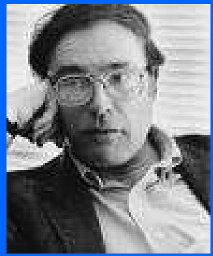
$$\sum_{n=0}^{\infty} B^n \frac{x^n}{n!} = e^{Bx}$$

$$e^{(B+1)x} - e^{Bx} \approx x$$

$$(B+1)^n - B^n \approx \delta_{1,n}$$

$$\sum_{k=0}^{n-1} \binom{n}{k} B_k = \delta_{1,n}$$

Where?



(1932-99)

- **The Classical Umbral Calculus**.
  (1994) *SIAM J. Math. Anal.*, G.-C. Rota, B. Taylor
- **The number of partition of a set**
  (1964) *Amer. Math. Monthly, G.C. Rota*
- **A selected survey of umbral calculus**
  (2000) Electron. J. Combin A. Di Bucchianico, D. Loeb
- **Umbral nature of Poisson random variables**
  (2001) *Algebraic Combinatorics and Computer Science: a tribute to G.C. Rota (Springer)* E. Di Nardo, D. Senato
- **An umbral setting for cumulants and factorial moments**
  (2006) *Europ. Jour. Comb.* E. Di Nardo, D. Senato

The classical umbral calculus

- $A=\{\alpha,\beta,\gamma,\dots\}$ *alphabet* whose elements are called *umbrae*
- R the field of real numbers
- A *linear functional* (evaluation) $E: R[A] \to R$ such that

  $i)\ E[1]=1;$

  $ii)\ E\left[\alpha^i \gamma^j \delta^k \cdots\right] = E\left[\alpha^i\right] E\left[\gamma^j\right] E\left[\delta^k\right] \cdots$ *uncorrelation property*

Special umbrae:

- *augmentation* $\Rightarrow$ ε such that $E\left[\varepsilon^k\right] = \delta_{0,k}, \forall k \in N$
- *unity* $\Rightarrow u$ such that $E\left[u^k\right] = 1, \forall k \in N$

A sequence $1, a_1, a_2, \ldots$ in R is said to be *umbrally represented* by an umbra α when

$$E\left[\alpha^i\right] = a_i, \quad i = 0, 1, 2, \ldots$$

MOMENTS

- *It has been argued that the notions of sample space and event are redundant, and that all of probability should be done in terms of random variables alone…How would one introduce probability in terms of random variables alone? This is a subject that has been thoroughly studied and a brief mention will suffice. One takes an ordered commutative algebra over the reals, and endows it with a positive linear functional E[X]. The elements of the algebra will be the random variables and the linear functional is the expectation of a random variable.*

G.C. Rota, *Twelve problems in probability no ones like to bring up – problem one: the algebra of probability*. (The Fubini Lectures,1998)

Umbral equivalence

$$\sum_{i=0}^{n}\binom{n}{i} a_i \, g_{n-i} = \sum_{i=0}^{n}\binom{n}{i} E\left[\alpha^i\right] E\left[\gamma^{n-i}\right] = E\left[(\alpha+\gamma)^n\right]$$

α and γ are *umbrally equivalent* when

$E[\alpha] = E[\gamma]$ in symbols $\alpha \approx \gamma$

$$\sum_{i=0}^{n}\binom{n}{i} a_i \, g_{n-i} \approx (\alpha+\gamma)^n$$

$$\sum_{i=0}^{n}\binom{n}{i} a_i \, a_{n-i} = E\left[(\alpha+\alpha')^n\right] \text{ to remind } E\left[(\alpha')^i\right] = a_i$$

Similarity equivalence

Two umbrae α and γ are said to be *similar* when

$$E\left[\alpha^n\right] = E\left[\gamma^n\right],\ n = 0,1,2,\ldots \quad \text{in symbols} \quad \alpha \equiv \gamma$$

$$\underbrace{\alpha + \alpha' + \cdots + \alpha''}_{m} \text{ where } \alpha, \alpha', \ldots, \alpha'' \text{ are distinct and similar}$$

$$m \cdot \alpha$$

Auxiliary umbrae – Saturated umbral calculus

$$\sum_{k_1 + k_2 + \cdots + k_m = n} \binom{n}{k_1, k_2, \ldots, k_m} a_{k_1} a_{k_2} \cdots a_{k_m} \simeq \Big(\underbrace{\alpha + \alpha' + \cdots + \alpha''}_{m}\Big)^n$$

$$\alpha^{k_1} (\alpha')^{k_2} \cdots (\alpha'')^{k_m}$$

Why?

General applicability of symmetric function algebra in statistical calculations involving moment (of moment) estimations

↓

The idea of AUGMENTED SYMMETRIC FUNCTIONS we believe to be ours

F.N. David, M.G. Kendall, D. E. Barton – Tables (1966) - Pearson

A FUNDAMENTAL RESULT IN MOMENT ESTIMATION

$$E\left[\sum \underbrace{X_i X_j \cdots}_{c_1} \underbrace{X_q^2 X_r^2 \cdots}_{c_2} \cdots \underbrace{X_u^m X_v^m \cdots}_{c_m}\right] = (n)_{\nu_\lambda} (\mu_1)^{c_1} (\mu_2)^{c_2} \cdots (\mu_m)^{c_m}$$

where $\lambda \equiv \left(1^{c_1}, 2^{c_2}, \ldots, m^{c_m}\right)$ is a *partition* of the integer $k \leq n$ in ν_λ parts,

that is $c_1 + 2\, c_2 + \cdots + m\, c_m = k$ and $c_1 + c_2 + \cdots + c_m = \nu_\lambda$

Why?

- **Algebraic complexity** → Umbral calculus

+

- **Amount of work required to reach it** → Symbolic computation

A FUNDAMENTAL RESULT IN MOMENT ESTIMATION

$$E\left[\sum \underbrace{X_i X_j \cdots}_{c_1} \underbrace{X_q^2 X_r^2 \cdots}_{c_2} \cdots \underbrace{X_u^m X_v^m \cdots}_{c_m}\right] = (n)_{\nu_\lambda} (\mu_1)^{c_1} (\mu_2)^{c_2} \cdots (\mu_m)^{c_m}$$

where $\lambda \equiv \left(1^{c_1}, 2^{c_2}, \ldots, m^{c_m}\right)$ is a *partition* of the integer $k \leq n$ in ν_λ parts,

that is $c_1 + 2\, c_2 + \cdots + m\, c_m = k$ and $c_1 + c_2 + \cdots + c_m = \nu_\lambda$

- **Algebraic complexity**

Umbral polynomial

A polynomial $p \in R$ is called an *umbral polynomial*.

$\alpha + \alpha' + \cdots + \alpha''$

$\chi\alpha + \chi'\alpha' + \cdots + \chi''\alpha''$

Special umbrae:

- *singleton* $\Rightarrow$ χ such that $E\left[\chi^k\right] = \delta_{1,k}, \; k = 1, 2, \ldots$

$$(\chi\alpha + \chi'\alpha' + \cdots + \chi''\alpha'')^2 = \sum \chi^2\alpha^2 + \sum (\chi\alpha)(\chi'\alpha')$$

E

$$\underbrace{(\chi\alpha + \chi'\alpha' + \cdots + \chi''\alpha'')^2}_{\left[n \boldsymbol{.} (\chi\alpha)\right]^2} \approx \sum (\chi\alpha)(\chi'\alpha') \approx \sum \alpha\,\alpha'$$

Umbral elementary symmetric polynomial

Umbral polynomial

$$\left[n\cdot(\chi\alpha)\right]^2\left[n\cdot\left(\chi\alpha^3\right)\right] \Leftarrow (\chi\alpha+\cdots+\chi'\alpha')^2\left(\chi\alpha^3+\cdots+\chi'(\alpha')^3\right)$$

$$=\left[\sum \chi^2\alpha^2+\sum(\chi\alpha)(\chi'\alpha')\right]\left(\sum\chi\alpha^3\right) \quad \approx \sum(\chi\alpha)(\chi'\alpha')\left(\chi''\alpha''^3\right)$$

Augmented symmetric polynomial

A fundamental result in Moment Estimation

If $\lambda \equiv \left(1^{c_1}, 2^{c_2}, \ldots, m^{c_m}\right) \vdash k \leq n$ in ν_λ parts, then

$$\left[n\cdot(\chi\alpha)\right]^{c_1}\left[n\cdot\left(\chi\alpha^2\right)\right]^{c_2}\cdots\left[n\cdot\left(\chi\alpha^m\right)\right]^{c_m} \approx (n)_{\nu_\lambda}\,\alpha_\lambda$$

where $\alpha_\lambda \equiv \alpha^{\bullet c_1}\left(\alpha'^2\right)^{\bullet c_2}\cdots\left(\alpha''^m\right)^{\bullet c_m}$ and $\alpha^{\bullet i} \equiv \underbrace{\alpha\,\alpha'\cdots\alpha''}_{i}$.

U-statistic

$$U = \frac{1}{(n)_k} \sum \Phi\left(X_{j_1}, X_{j_2}, \ldots, X_{j_k}\right)$$

where $X_1, X_2, \ldots, X_n$ are n i.r.v.'s, and the sum ranges in the set of all permutations $(j_1, j_2, \ldots, j_k)$ of k integers with $1 \le j_i \le n$.

When $X_1, X_2, \ldots, X_n$ are i.d., U results a *symmetric polynomial*

By virtue of the fundamental theorem on symmetric polynomials, such an U statistic can be expressed as a polynomial in elementary symmetric polynomials and so

$$\left[n \cdot (\chi\alpha)\right]^{c_1} \left[n \cdot \left(\chi\alpha^2\right)\right]^{c_2} \cdots \left[n \cdot \left(\chi\alpha^m\right)\right]^{c_m}$$

k-statistic

A unyfing framework for k-statistics, polykays and their generalizations. (2008) *Bernoulli*

E. Di Nardo, G. Guarino, D. Senato

$$(\chi \cdot \alpha)^i \simeq \sum_{\lambda \vdash i} (\chi)_{\nu_\lambda} \, d_\lambda \alpha_\lambda$$

$$\simeq \sum_{\lambda \vdash i} \frac{(\chi)_{\nu_\lambda}}{(n)_{\nu_\lambda}} d_\lambda \left[n.(\chi\alpha) \right] \left[n.\left(\chi\alpha^2\right) \right]^{c_2} \cdots \left[n.\left(\chi\alpha^m\right) \right]^{c_m}$$

A fundamental formula

$$(n \cdot \alpha)^i \simeq \sum_{\lambda \vdash i} (n)_{\nu_\lambda} \, d_\lambda \alpha_\lambda$$

$$E\left[(\chi)_k \right] = (-1)^k (k-1)!$$

$$d_\lambda = \frac{i!}{(1!)^{c_1} (2!)^{c_2} \cdots (m!)^{c_m} c_1! c_2! \cdots c_m!}$$

Generatingfunctionology

(Herbert S. Wilf 1990)

- **Amount of work required to reach it**

AUGMENTED SYMMETRIC FUNCTIONS —In terms of→ POWER SUMS

$$n \cdot \alpha^r \equiv \underbrace{\alpha^r + \alpha'^r + \cdots + \alpha''^r}_{n}$$

- **Symbolic computation for statistical inference** (2000) Oxford Statistical Science Series, D.F. Andrews, J.E. Stafford

- **MathStatica: a symbolic approach to computational mathematical statistics** (2008) C. Rose (next speaker)

- **Symbolic computation of moments of sampling distributions** (2008) *Comp. Stat. Data Analysis* E. Di Nardo, G. Guarino, D. Senato

Main formulae

$$(n\cdot\mu)_M \approx \sum_{S_\pi}\left[n\cdot(\chi\mu)\right]_{S_\pi}$$

$$\left[n\cdot(\chi\mu)\right]_M \approx \sum_{S_\pi}(\chi\cdot\chi)^{\bullet S_\pi}\left[n\cdot\mu\right]_{S_\pi}$$

where

$$M = \left\{\underbrace{\mu_1,\ldots,\mu_1}_{f(\mu_1)},\ldots,\underbrace{\mu_k,\ldots,\mu_k}_{f(\mu_k)}\right\}$$ a multiset, S_π is a subdivision (partition) of M

Main formulae

$$(n \cdot \mu)_M \approx \sum_{S_\pi} \left[n \cdot (\chi \mu) \right]_{S_\pi}$$

$$\left[n \cdot (\chi \mu) \right]_M \approx \sum_{S_\pi} (\chi \cdot \chi)^{\bullet S_\pi} \left[n \cdot \mu \right]_{S_\pi}$$

Example:

$$M = \Big\{ \underbrace{\alpha, \ldots, \alpha,}_{i} \Big\} \text{ then } S_\pi = \Big\{ \underbrace{\{\alpha\}, \ldots, \{\alpha\}}_{c_1}, \underbrace{\{\alpha, \alpha\}, \ldots, \{\alpha, \alpha\}}_{c_2}, \ldots \Big\}$$

$$(n \cdot \mu)_M \approx (n \cdot \alpha)^i$$

$$\left[n \cdot (\chi \mu) \right]_{S_\pi} \approx \left[n \cdot (\chi \alpha) \right]^{c_1} \left[n \cdot \left(\chi \alpha^2 \right) \right]^{c_2} \cdots$$

Main formulae

$$(n\cdot\mu)_M \approx \sum_{S_\pi}\left[n\cdot(\chi\mu)\right]_{S_\pi}$$

$$\left[n\cdot(\chi\mu)\right]_M \approx \sum_{S_\pi}(\chi\cdot\chi)^{\bullet S_\pi}\left[n\cdot\mu\right]_{S_\pi}$$

Example:

$$(n\cdot\mu_1)^2(n\cdot\mu_2)$$

$$M=\{\mu_1,\mu_1,\mu_2\} \text{ then } S_\pi=\begin{cases}\{\{\mu_1\},\{\mu_1,\mu_2\}\} \Rightarrow \sum X_i X_j Y_j \\ \{\{\mu_2\},\{\mu_1,\mu_1\}\} \Rightarrow \sum X_i^2 Y_j \\ \{\{\mu_1\},\{\mu_2\},\{\mu_1\}\} \Rightarrow \sum X_i X_j Y_k \\ \{\{\mu_1,\mu_1,\mu_2\}\} \Rightarrow \sum X_i^2 Y_i\end{cases}$$

$$E\left(\sum_{i=1}^{n} X_i\right)^2\left(\sum_{i=1}^{n} Y_i\right) = 2(n)_2\,E[X]E[XY]+(n)_2\,E\left[X^2\right]E[Y]+\cdots$$

Main formulae

$$(n.\mu)_M \approx \sum_{S_\pi} \left[n.(\chi\mu) \right]_{S_\pi}$$

$$\left[n.(\chi\mu) \right]_M \approx \sum_{S_\pi} (\chi.\chi)^{\bullet S_\pi} \left[n.\mu \right]_{S_\pi}$$

Example:

$$\left(\sum X_i^2 X_j\right)\left(\sum X_i^2 Y_i\right)^2 \longrightarrow \sum X_{i_1}^{k_1} Y_{i_1}^{l_1} X_{i_2}^{k_2} Y_{i_2}^{l_2} \cdots$$

$$M = \left\{ \chi_1\mu_1^2, \chi_1\mu_1, \chi_2\mu_1^2\mu_2, \chi_3\mu_1^2\mu_2 \right\}$$

The subdivision $S_\pi = \left\{ \underbrace{\left\{ \chi_1\mu_1^2, \chi_1\mu_1 \right\}}, \left\{ \chi_2\mu_1^2\mu_2 \right\}, \left\{ \chi_3\mu_1^2\mu_2 \right\} \right\}$ gives zero contribution due to the singleton umbra

Special results

It is possible to build fast algorithms by forfeiting the elegant idea to produce one algorithm for the whole subject.

K-statistics ← **Cumulants of compound Poisson random variables**

~~AUGMENTED POLYNOMIALS~~ **Exponential polynomials**

$$(\chi \cdot \alpha)^i \approx \sum_{\lambda \vdash i} d_\lambda p_\lambda \left(\frac{\chi \cdot \chi}{n \cdot \chi} \right) (n \cdot \alpha)^{c_1} \left(n \cdot \alpha^2\right)^{c_2} \cdots$$

where $p_\lambda(y) = \left[p_1(y)\right]^{r_1} \left[p_2(y)\right]^{r_2} \cdots$

and $p_n(y) = \sum_{i=1}^{n} y^k S(n,k)(-1)^{k-1}(k-1)!$

- PC Pentium(R)4 Intel(R) CPU 2.08 Ghz 512MB Ram Maple 10.0 Mathematica 4.2

- Mac OSX CPU 2.8Ghz Mathematica 6.0.2 Private communi-cation, C.Rose

K-statistics	A&S	Mathstatica 1	Umbral	Mathstatica 2
k_5	0.06	0.01	0.01	0.008
k_7	0.31	0.02	0.01	0.017
k_9	1.44	0.04	0.01	0.039
k_{11}	8.36	0.14	0.01	0.084
k_{14}	396.39	0.64	0.02	0.329
k_{16}	57982.4	2.63	0.08	0.917
k_{18}	-	6.90	0.16	2.080
k_{20}	-	25.15	0.33	9.363
k_{22}	-	81.70	0.80	32.114
k_{24}	-	359.40	1.62	
k_{26}	-	1581.05	2.51	
k_{28}	-	6505.45	4.83	

A new method for fast computing unbiased estimators of cumulants (2008) *Statistics and Computing* E.Di Nardo, G. Guarino, D. Senato

- PC Pentium(R)4 Intel(R) CPU 2.08 Ghz 512MB Ram Maple 10.0 Mathematica 4.2

- Mac OSX CPU 2.8Ghz Mathematica 6.0.2 Private communication, C.Rose

Multivariate Polykays	A&S	Umbral	Mathstatica 2
$k_{1\ 1;1\ 1}$	0.05	0.01	0.006
$k_{2\ 1;1\ 1}$	0.20	0.01	0.014
$k_{2\ 2;\ 1\ 1}$	1.22	0.03	0.038
$k_{2\ 2;2\ 1}$	6.30	0.08	0.085
$k_{2\ 2;2\ 2}$	33.75	0.14	0.204
$k_{2\ 1;2\ 1;2\ 1}$	78.94	0.22	0.227
$k_{2\ 2;1\ 1;1\ 1}$	30.01	0.14	0.154
$k_{2\ 2;2\ 1;2\ 1}$	398.42	0.55	0.928
$k_{2\ 2;2\ 2;1\ 1}$	464.45	0.61	1.063
$k_{2\ 2;2\ 2;2\ 1}$	1387.00	1.25	2.622
$k_{2\ 2;2\ 2;2\ 2}$	3787.41	2.91	6.402

Why k-statistics?

Cumulants as non-Gaussian qualifiers *Ferreira,P.; Magueijo, J. Silk, J.* Phys.Rev. D (1997)

Wavelet analysis and the detection of non-Gaussianity in the cosmic microwave background. *Hobson, M. P.; Jones,*

A. *W.; Lasenby, A. N. Monthly Notices of the Royal Astronomical*
B. *Society, (1999)*

The higher-order k-statistics are listed by Stuart & Ord (1994) up to r = 8 but are extremely lengthy to write out in full.

Cumulant Analysis in Fluorescence Fluctuation Spectroscopy.
Muller, J.D. Biophys J. (2004)

Higher Order Cumulants of Random Vectors and Applications to Statistical Inference and Time Series. *Rao J., Subba T., Terdik G*. Sankhya (2006)

Concluding remarks

On the basis of this result, the umbral calculus can be interpreted as a calculus of measures on Poisson algebras, generalizing compound Poisson processes. **The umbral calculus S. Roman, G.C. Rota (1978) Adv. Math.**

- Unbiased estimators of population moments *H-statistics*
- Population moments of sample moments
- L-moments and trimmed L-moments have been noticed as appealing alternatives to conventional moments.

Karvanen, J., 2006. Estimation of quantile mixtures via L-moments and trimmed L-moments. Comput. Statist. Data Anal., 51, 947-959.

THANK YOU FOR YOUR ATTENTION