Kernel families of probability measures

Saskatoon, October 21, 2011

Abstract

The talk will compare two families of probability measures: exponential, and Cauchy-Stjelties families. The exponential families have long and rich history, starting with Fisher in 1934. The Cauchy-Stjelties families show similarities or parallels to exponential families but have been studied much less.

Outline (of some other talk)

Kernel families

Natural exponential families

- Parametrization by the mean*
- Variance function*
- Reproductive property*
- Exponential families of stochastic processes
- Martingale characterizations of randomization laws
- Multivariate exponential families

Cauchy-Stjelties families

- Parametrization by the mean*
- Variance function*
- Reproductive property*
- Extending domain of means
- Iterated CSK families

q-exponential families

Kernel families

For real x, θ, suppose that $k(x, \theta)$ is integrable with respect to a positive σ-finite measure $\mu(dx)$, and is positive for all $\theta \in \Theta$ and all x in the support of μ. Denote

$$L(\theta) = \int k(x, \theta)\mu(dx).$$

Definition (Wesolowski (1999))

The kernel family is a family of probability measures on Borel subsets of $\mathbb{R}$ defined as

$$\mathcal{K} = \mathcal{K}(\mu, k) = \left\{ P_\theta(dx) = \tfrac{1}{L(\theta)} k(x, \theta)\mu(dx) : \theta \in \Theta \right\}$$

Jacek Wesołowski 1999

KERNEL FAMILIES

Let μ be a positive measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, and $k : \mathbf{R}^2 \to \mathbf{R}_+$ a measurable function, wkich will play the role of the kernel. Denote

$$G_\mu(\theta) = \int_{\mathbf{R}} k(x,\theta)\,\mu(dx),$$

for such θ that $\int_{\mathbf{R}} k(x,\theta)\,\mu dx < \infty$. Denote additionally

$$\Theta = Int\{\theta \in \mathbf{R} : \int_{\mathbf{R}} k(x,\theta)\,\mu(dx) < \infty\}.$$

The natural exponential family is an example of a kernel family with $k(x,\theta) = e^{\theta x}$. Consider two new special cases:

GAMMA KERNEL FAMILY

Example (Natural exponential family)

Suppose $\int e^{\theta x}\mu(dx) < \infty$ for some θ. Let $k(x,\theta) = e^{\theta x}$.
The *natural exponential family* (NEF) generated by μ is:

$$\mathcal{K}(\mu) = \left\{ P_\theta(dx) = \tfrac{1}{L(\theta)} e^{\theta x}\mu(dx) : \theta \in \Theta \right\}$$

where $L(\theta) = \int e^{\theta x}\mu(dx)$.

- It is convenient to take open interval $\Theta = (\theta_-, \theta_+)$ even if the Laplace transform $L(\theta)$ is defined at the end-points.
- $\mathcal{K}(\mu) = \mathcal{K}(P_{\theta_0})$ so WLOG can assume that μ is a probability measure.

(pre) History
(of more general exponential families)

From http://en.wikipedia.org/wiki/ and other internet sources

- Fisher, R. (1934) *Two new properties of mathematical likelihood*, Proc. Roy. Soc.
- Pitman, E.; Wishart, J. (1936). *Sufficient statistics and intrinsic accuracy*. Mathematical Proceedings of the Cambridge Philosophical Society
- Darmois, G. (1935). *Sur les lois de probabilites a estimation exhaustive*. C.R. Acad. Sci. Paris
- Koopman, B (1936). *On distribution admitting a sufficient statistic*. Transactions of the American Mathematical Society

Good statistical properties

Under suitable assumptions on μ,

- Sample average is a sufficient statistics for θ
- Maximal likelihood estimator of θ is an "explicit" function of the sample mean.
- Maximal likelihood estimator of θ is asymptotically normal.
- likelihood ratio test is approximately chi-square

- If θ is random, explicit "conjugate priors" have linear regression of mean onto observation Diaconis-Ylvisaker (1979)
- If X has law from the Morris class and θ is random with a "conjugate prior", then the marginal law of X is known.

None of these questions are the topic of this talk. Morris class will appear below.

Example (Cauchy-Stjelties Kernel family)

Consider kernel

$$k(x,\theta) = \frac{1}{1 - x\theta}$$

and a *probability measure* μ with support in $(-\infty, B)$.
The Cauchy-Stjelties kernel (CSK) family is

$$\mathcal{K} = \left\{ P_\theta(dx) = \frac{1}{L(\theta)(1-\theta x)}\mu(dx) : \theta \in \Theta \right\}$$

where $L(\theta) = \int (1 - x\theta)^{-1}\mu(dx)$.

- It is convenient to analyze one-sided CSK families with $\theta \in \Theta = (0, \theta_+)$,
- We take $B = \sup\{b : \mu(b, \infty) > 0\}$ and $\theta_+ = 1/B$ if $B > 0$ (with $\theta_+ = \infty$ if $B \leq 0$).

NEF versus CSK families

The talk will switch between NEF and CSK families generated by μ

$$\mathcal{K}(\mu) = \{P_\theta(dx) : \theta \in \Theta\}$$

- NEF :

$$P_\theta(dx) \propto e^{\theta x}\mu(dx)$$

- CSK:

$$P_\theta(dx) \propto \frac{1}{1-\theta x}\mu(dx)$$

A specific example of NEF

Noncanonical parameterization

Let μ be a counting measure on $\{0, 1, \dots\}$ given by $\mu(\{j\}) = 1$

- $L(\theta) = \int e^{\theta x} \mu(dx) = \sum_{j=0}^{\infty} e^{j\theta} = 1/(1 - e^{\theta})$.
- $\theta \in \Theta = (-\infty, 0)$
- $$\mathcal{K}(\mu) = \left\{ P_\theta(dx) : P_\theta(\{j\}) = (1 - e^{\theta}) e^{\theta j},\ \theta < 0 \right\}$$
- More standard parametrization: $p = e^{\theta}$

$$\mathcal{K}(\mu) = \{ Q_p(dx) : Q_p(\{j\}) = (1 - p) p^j,\ 0 < p < 1 \}$$

- $\mathcal{K}(\mu)$ is the family of (all) geometric probability laws.
- $\mathcal{K}(Q_{1/2})$ is the same family!

Skip second example

Another example of NEF

Again in noncanonical parametrization

Let μ be a discrete measure concentrated on $\{0, 1, \dots\}$ given by

$$\mu(\{j\}) = 1/j!$$

- Then $L(\theta) = \int e^{\theta x} \mu(dx) = \sum_{j=0}^{\infty} e^{j\theta}/j! = \exp(e^{\theta})$.
- $\mathcal{K}(\mu) = \left\{ P_\theta(dx) : P_\theta(\{j\}) = \exp(-e^{\theta}) e^{\theta j}/j!,\ \theta \in (-\infty, \infty) \right\}$
- More standard parametrization: $\lambda = e^{\theta} > 0$
- $\mathcal{K}(\mu) = \{ Q_\lambda(dx) : Q_\lambda(\{j\}) = e^{-\lambda} \lambda^j / j!,\ \lambda > 0 \}$
- Poisson family parametrized by $\lambda > 0$.

Parametrization by the mean

$$\mathcal{K}(\mu) = \left\{ P_\theta(dx) = e^{\theta x - \kappa(\theta)} \mu(dx) : \theta \in \Theta \right\}$$

$$\kappa(\theta) = \log L(\theta) = \log \int e^{\theta x} \mu(dx)$$

- The mean is $m(\theta) = \int x P_\theta(dx) = \kappa'(\theta) = L'/L$
- The variance (of P_θ) is $\kappa''(\theta) > 0$ for non-degenerate μ.

Parametrization by the mean

$$m(\theta) = \int x P_\theta(dx) = \kappa'(\theta)$$

- For non-degenerate measure μ, function $\theta \mapsto m(\theta)$ is strictly increasing and has inverse inverse $\theta = \psi(m)$.
- κ' maps (θ_-, θ_+) onto (m_-, m_+), "the domain of means".
- Parameterization by the mean:

$$\mathcal{K}(\mu) = \{Q_m(dx) : m \in (m_-, m_+)\}$$

where $Q_m(dx) = P_{\psi(m)}(dx)$, i.e. $Q_{m(\theta)}(dx) = P_\theta(dx)$, i.e. $Q_{\kappa'(\theta)}(dx) = P_\theta(dx)$.

Remark

$\theta \mapsto \int \frac{x}{L(\theta)(1-\theta x)}\mu(dx)$ is also strictly increasing. So CSK families can also be parametrized by the mean.

Variance function of a NEF

Morris (1982)

$$V(m) = \int (x-m)^2 Q_m(dx)$$

- Variance function $V(m)$ (together with the domain of means $m \in (m_-, m_+)$) determines NEF uniquely. So $\mathcal{K}(\mu) = \mathcal{K}(V)$.
- No simple formula for $Q_m(dx)$ in terms of $V(m)$. However, Ismail-May (1978) give

$$\frac{\partial}{\partial m} Q_m(dx) = \frac{x-m}{V(m)} Q_m(dx)$$

- Which functions $V(m)$ are variance functions of some NEF? Jörgensen's criterion.
- Continuity Mora (1987): $\lim_{k\to\infty} V_k(m) = V(m)$ uniformly on a closure of some open interval, then V is the variance function and $Q_m^{(k)} \in \mathcal{K}(V_k)$ converge weakly to $Q_m \in \mathcal{K}(V)$ as $k \to \infty$.

Morris class

Theorem (Morris (1982), Ismail-May (1978))

Suppose $b \geq -1$. The NEF with the variance function $V(m) = 1 + am + bm^2$ consists of

1. *the Gaussian laws of variance 1 if $a = b = 0$;*
2. *the Poisson type laws if $b = 0$ and $a \neq 0$;*
3. *the Pascal (negative binomial) type laws if $b > 0$ and $a^2 > 4b$;*
4. *the Gamma type law if $b > 0$ and $a^2 = 4b$;*
5. *the hyperbolic secant type laws if $b > 0$ and $a^2 < 4b$;*
6. *the binomial type laws if $b = -1/n$ for some $n \in \mathbb{N}$;*

- This covers all possible quadratic variance functions normalized to $V(0) = 1$.
- $Q_0(dx)$ is standardized to have mean zero variance 1
- Letac-Mora (1990): cubic $V(m)$
- Various other classes Kokonendji, Letac, ...

Variance function of a CSK family

$m_0 = \int x\mu(dx) \in [-\infty, B)$.

$$\mathcal{K}(\mu) = \left\{ \frac{\mu(dx)}{L(\theta)(1-\theta x)} : \theta \in \Theta \right\} = \{Q_m(dx) : m \in (m_0, m_+)\}$$

$$m = m(\theta) = \frac{L(\theta) - 1}{\theta L(\theta)} = D_0(L)/L$$

If μ has the first moment then $V(m) = \int (x-m)^2 Q_m(dx)$ is well defined.

- Variance function $V(m)$ (together with $m_0 \in \mathbb{R}$) determines μ and hence the CSK family. So $\mathcal{K}(\mu) = \mathcal{K}(V)$.
- Explicitly, $Q_m(dx) = \frac{V(m)}{V(m)+(m-m_0)(m-x)}\mu(dx)$
- Which functions $V(m)$ are variance functions of some CSK?
- Continuity W.B. (2009): yes, at least for compactly supported measures. (The statement is more technical with fixed $m_0 = \int x\mu_k(dx)$.)

Theorem (Theorem 4.2, WB.-Ismail (2005))

Suppose $b \geq -1$, $m_0 = 0$. The CSK family with the variance function $V(m) = 1 + am + bm^2$ has generating measure μ:

1. *the Wigner's semicircle (free Gaussian) law if $a = b = 0$; see [Voiculescu, 2000, Section 2.5];*
2. *the Marchenko-Pastur (free Poisson) type law if $b = 0$ and $a \neq 0$; see [Voiculescu, 2000, Section 2.7];*
3. *the free Pascal (negative binomial) type law if $b > 0$ and $a^2 > 4b$; see [Saitoh and Yoshida, 2001, Example 3.6];*
4. *the free Gamma type law if $b > 0$ and $a^2 = 4b$; see [Bożejko and Bryc, 2006, Proposition 3.6];*
5. *the free analog of hyperbolic type law if $b > 0$ and $a^2 < 4b$; see [Anshelevich, 2003, Theorem 4];*
6. *the free binomial type law (Kesten law, McKay law) if $-1 \leq b < 0$; see [Saitoh and Yoshida, 2001, Example 3.4].*

(Standardized to have mean zero variance 1) End now

Reproductive property for NEF

Theorem (Jörgensen (1997))

If ν is a probability measure in NEF with variance function $V(m)$, then for $n \in \mathbb{N}$ the law of the sample mean, $\nu_n(U) := (\nu * \nu * \cdots * \nu)(nU)$, is in NEF with variance function

$$V_n(m) = \frac{V(m)}{n}.$$

Reproductive property for CSK families

For $r \geq 1$, let $\mu^{\boxplus r}$ denote the r-fold free additive convolution of μ with itself.

Theorem (WB-Ismail (2005))

If a function $V(m)$ analytic at m_0 is a variance function of a CSK family generated by a compactly supported probability measure ν with $m_0 = \int x \nu(dx)$, then for each $r \geq 1$ there exists a neighborhood of m_0 such that $V(m)/r$ is the variance function of the CKS family generated by

$$\nu_r(U) := \nu^{\boxplus r}(rU).$$

In contrast to NEF, the neighborhood of m_0 where $m \mapsto V(m)/r$ is a variance function may vary with r.

▸ Approximation operators ▸ q-exponential families

Summary

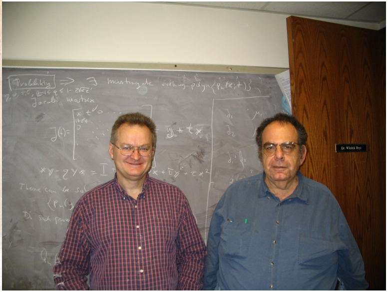
Kernels $e^{\theta x}$ and $1/(1-\theta x)$ generate NEF and CSK families

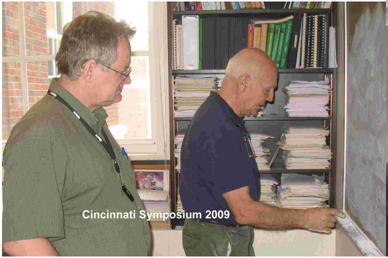
Similarities

- parameterization by the mean
- Quadratic variance functions determine interesting laws
- For NEF, convolution affects variance function in a similar way as additive free convolution for CSK

Differences

- The generating measure of a NEF is not unique.
- A CSK family in parameterization by the mean may be well defined beyond the "domain of means"
- For CSK family, the variance function may be undefined. Instead of the variance function [Bryc and Hassairi, 2011] look at the "pseudo-variance" function $m \mapsto mV(m)/(m-m_0)$ which is well defined for more measures μ.





Thank you

Anshelevich, M. (2003).
Free martingale polynomials.
Journal of Functional Analysis, 201:228–261.
arXiv:math.CO/0112194.

Bożejko, M. and Bryc, W. (2006).
On a class of free Lévy laws related to a regression problem.
J. Funct. Anal., 236:59–77.
arxiv.org/abs/math.OA/0410601.

Bryc, W. (2009).
Free exponential families as kernel families.
Demonstr. Math., XLII(3):657–672.
arxiv.org:math.PR:0601273.

Bryc, W. and Hassairi, A. (2011).
One-sided Cauchy-Stieltjes kernel families.
Journ. Theoret. Probab., 24(2):577–594.
arxiv.org/abs/0906.4073.

Bryc, W. and Ismail, M. (2005).

Approximation operators, exponential, and q-exponential families.
Preprint. arxiv.org/abs/math.ST/0512224.

Bryc, W. and Wesołowski, J. (2005).
Conditional moments of q-Meixner processes.
Probab. Theory Related Fields Fields, 131:415–441.
arxiv.org/abs/math.PR/0403016.

Saitoh, N. and Yoshida, H. (2001).
The infinite divisibility and orthogonal polynomials with a constant recursion formula in free probability theory.
Probab. Math. Statist., 21(1):159–170.

Voiculescu, D. (2000).
Lectures on free probability theory.
In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, volume 1738 of *Lecture Notes in Math.*, pages 279–349. Springer, Berlin.

Thank you

Back

U. Küchler, Exponential Families of Stochastic Processes.
Springer, 1997.

Nekrutkin (200?)

Definition

Approximation operators and exponential families

$$\lim_{\lambda \to \infty} S_\lambda(f)(m) = f(m)$$

C. P. May (1976) and Ismail and May(1978) consider *exponential type operators* as positive operator

$$S_\lambda(f)(m) = \int_{\mathbb{R}} W_\lambda(m, x) f(x) dx,$$

where W_λ is a generalized function satisfying

$$\frac{\partial W_\lambda(m, x)}{\partial m} = \lambda \frac{x - m}{V(m)} W_\lambda(m, x), \quad \lambda > 0, \tag{1}$$

$$\int_{\mathbb{R}} W_\lambda(m, x)\, dx = 1. \tag{2}$$

End now

Since

$$\int_{\mathbb{R}} W_\lambda(m,x)\, x\, dx = m, \qquad \int_{\mathbb{R}} W_\lambda(m,x)\,(x-m)^2 dx = \frac{V(m)}{\lambda}$$

so m and $V(m)/\lambda$ are the mean and variance of $W_\lambda(m,x)$, respectively.

Measures in a NEF with variance function $V(m)/\lambda$ satisfy the differential equation and give rise to approximation operators of exponential type.

End now

q-exponential families

Recall that for $-1 < q < 1$ the q-differentiation operator is

$$(D_{q,m}f)(m) := \frac{f(m) - f(qm)}{m - qm} \quad \text{for } m \neq 0.$$

Suppose $w(m, u)$ satisfies

$$D_{q,m}w(m, x) = w(m, x)\, \frac{x - m}{V(m)}.$$

This equivalent to

$$w(m, x) = \frac{w(mq, x)}{1 + m(1 - q)(m - x)/V(m)}. \tag{3}$$

which has the solution

$$w(m, x) = \prod_{n=0}^{\infty} \frac{V(q^n m)}{V(q^n m) + q^n m(1 - q)(q^n m - x)}, \tag{4}$$

provided that the infinite products converge.

Definition

A family of probability measures

$$\mathcal{F}(V) = \{w(m,x)\mu(dx):\ m \in (A,B)\}$$

is a q-exponential family with the variance function V if

1. μ is compactly supported,
2. $0 \in (A,B)$ and $\lim_{t\to 0} w(t,x) = w(0,x) \equiv 1$ for all $x \in \operatorname{supp}(\mu)$,
3. $V > 0$ on (A,B), $V(0) = 1$, and

$$D_{q,m} w(m,x) = w(m,x)\,\frac{u-m}{V(m)}$$

for all $m \neq 0$.

One can check that

$$\int x w(m,x)\mu(dx) = m,\ \int (x-m)^2 w(m,x)\mu(dx) = V(m).$$

Quadratic variance functions determine q-exponential families uniquely.

Theorem ([Bryc and Ismail, 2005])

If $\mathcal{F}(V)$ is a q-exponential family with the variance function

$$V(m) = 1 + am + bm^2$$

and $b > -1 + \max\{q, 0\}$ then

$$w(m,x) = \prod_{k=0}^{\infty} \frac{1 + amq^k + bm^2q^{2k}}{1 + (a - (1-q)x)mq^k + (b + 1 - q)m^2q^{2k}} \quad (5)$$

and $\mu(dx)$ is a uniquely determined probability measure with the absolutely continuous part supported on the interval $\frac{a}{1-q} - \frac{2\sqrt{b+1-q}}{1-q} < x < \frac{a}{1-q} + \frac{2\sqrt{b+1-q}}{1-q}$ and no discrete part if $a^2 < 4b$

For $b \geq 0$ the above μ appears in [Bryc and Wesołowski, 2005] in connection to a quadratic regression problem.

◂ End now