New Dirichlet Mean Identities

Lancelot James (lancelot@ust.hk)

Hong Kong University of Science and Technology

Isaac Newton Institute, August 10, 2007

Origins

CIFARELLI, D. M. and REGAZZINI, E. (1979). Considerazioni generali sull'impostazione bayesiana di problemi non parametrici. Le medie associative nel contesto del processo aleatorio di Dirichlet I, II. *Riv. Mat. Sci. Econom. Social* **2**, 39–52.

Later there is the important 1990 Annals of Statistics paper.

1. It is a cornerstone work in Bayesian NP and I believe one of the most important contributions in the theory, and indeed application, of random processes.
2. Why? Dirichlet means arise everywhere.

Some things to keep in mind

- Hard and very technical stuff
- But Cifarelli and Regazzini already did the hard work so we do not have to
- Our Task is really how to figure out how to use their results.
- Not so much on how to re-prove them.
- For some recent non-trivial applications look me up on the math Arxiv
- I never really do hard things. This is not a Joke

Theme

Prendere due piccioni con una fava

Dirichlet process

Let X be a non-negative random variable with cumulative distribution function F_X. Note furthermore for a measurable set C, we use the notation $F_X(C)$ to mean the probability that X is in C. One may define a Dirichlet process random probability measure, say P_θ, on $[0,\infty)$ with total mass parameter θ and *prior* parameter F_X, via its finite dimensional distribution as follows; for any disjoint partition on $[0,\infty)$, say $(C_1,\ldots,C_k)$, the distribution of the random vector. $(P_\theta(C_1),\ldots,P_\theta(C_k))$ is a k-variate Dirichet distribution with parameters $(\theta F_X(C_1),\ldots,\theta F_X(C_k))$.

Hence for each C,

$$P_\theta(C) = \int_0^\infty \mathbb{I}(x \in C) P_\theta(dx)$$

has a beta distribution with parameters $(\theta F_X(C), \theta(1 - F_X(C)))$.

Equivalently setting $\theta F_X(C_i) = \theta_i$ for $i = 1, \ldots, k$,

$$(P_\theta(C_1), \ldots, P_\theta(C_k)) \stackrel{d}{=} \left(\frac{G_{\theta_i}}{G_\theta}; i = 1, \ldots, k \right)$$

where (G_{θ_i}) are independent random variables with gamma$(\theta_i, 1)$ distributions and $G_\theta = G_{\theta_1} + \cdots + G_{\theta_k}$ has a gamma$(\theta, 1)$ distribution.

This means that one can define the Dirichlet process via the normalization of an independent increment gamma process on $[0, \infty)$, say $\gamma_\theta(\cdot)$, as

$$P_\theta(\cdot) = \frac{\gamma_\theta(\cdot)}{\gamma_\theta([0, \infty))}$$

where $\gamma_\theta(C_i) \stackrel{d}{=} G_{\theta_i}$ and whose almost surely finite total random mass is $\gamma_\theta([0, \infty)) \stackrel{d}{=} G_\theta$.

A very important aspect of this construction is the fact that G_θ is independent of P_θ, and hence any functional

$$P_\theta(g) = \int_0^\infty g(x) P_\theta(dx)$$

Furthermore $\theta > 0$,

$$G_\theta P_\theta(g) \stackrel{d}{=} \int_0^\infty g(x)\gamma_\theta(dx) = \gamma_\theta(g)$$

See for instance Lijoi and Regazzini- AOP

Now,

$$\mathbb{E}[\mathrm{e}^{-\lambda G_\theta P_\theta(g)}] = \mathbb{E}[(1 + \lambda P_\theta(g))^{-\theta}]$$

- The Laplace transform of a random variable representable as $\gamma_\theta(g)$ is equivalent to the Cauchy-Stieltjes transform of order θ of a random variable representable as $P_\theta(g)$

Something to keep in mind

Lets think about random variables on the real line rather than random processes in some Banach space

A more general statement: Psychologically easier to prove???!!!

Let M be a positive random variable independent of G_θ, and define the random variable $R_\theta \overset{d}{=} G_\theta M$ then from your first course in probability

$$\mathbb{E}[\mathrm{e}^{-\lambda R_\theta}] = \mathbb{E}[(1+\lambda M)^{-\theta}]$$

Question

- Are the results on the last two slides equivalent?

Answer-Questions

- $\gamma_\theta(g)$ is always infinitely divisible
- Scale mixtures of gamma random variables are not always infinitely divisible.
- Scale mixtures of gamma random variables of index $0 < \theta \leq 1$ are always infinitely divisible.
- If R_θ is infinitely divisible is it true then?

Classical Question

Recall that

$$\mathbb{E}[\mathrm{e}^{-\lambda R_\theta}] = \mathbb{E}[(1+\lambda M)^{-\theta}]$$

- In general :How to find the density of R_θ?
- In general :How to find the density of M?

Classical but painful answer

- Utilize Classical Inversion formula
- You have to deal with the Analysis of Complex functions
- No simpler answer in general

- Cifarelli and Regazzini provide answers for $M \stackrel{d}{=} P_\theta(g)$

My Solutions

- Call up my Italian friends!!
- (Lijoi and Prünster)
- Build on existing results

Important?

- $\gamma_\theta(g)$ constitute a large class of Infinitely divisible random variables.
- Id random variables generate EVERY positive Lévy process
- Levy processes are now the building blocks in many diverse fields, physics, finance, genetics, machine learning.
- All these models can be treated from a Bayesian NP point of view
- Levy processes are connected with special functions
- One wants to get the laws of certain functionals, BUT this is hard
- Simple example-Finance: One may want to use a Barndorff-Nielsen OU SV process for option pricing without resorting to a series representation.

1. Many important quantities can be represented as $P_\theta(g)$
2. Curiously James, Lijoi and Pruenster[JLP], show that for $\theta > 0$ every linear functional of a two parameter (α, θ) Poisson Dirichlet random probability measure can be represented as a Dirichlet mean of order θ.

Define a Dirichlet mean of order θ indexed by F_X as

$$M_\theta(F_X) = \int_0^\infty x P_\theta(dx)$$

Say that

$$T_\theta \stackrel{d}{=} G_\theta M_\theta(F_X)$$

is a GGC(θ, F_X) random variable which satisfies

$$\mathbb{E}[\mathrm{e}^{-\lambda T_\theta}] = \mathbb{E}[(1 + \lambda M_\theta(F_X))^{-\theta}] = \mathrm{e}^{-\theta \psi_{F_X}(\lambda)}$$

where

$$\psi_{F_X}(\lambda) = \int_0^\infty \log(1 + \lambda x) F_X(dx) = \mathbb{E}[\log(1 + \lambda X)].$$

Key expressions for density

Cifarelli and Regazzini(1990)

1. $$\Phi_{F_X}(t) = \int_0^\infty \log(|t-x|)I(t \neq x)F_X(dx) = \mathbb{E}[\log(|t-X|)\mathbb{I}(t \neq X)]$$

2. furthermore, define,

$$\Delta_\theta(t|F_X) = \frac{1}{\pi}\sin(\pi\theta F_X(t))\mathrm{e}^{-\theta\Phi_{F_X}(t)}.$$

Then from [CR]

1. The cdf of $M_\theta(F_X)$ for all $\theta > 0$ is expressible as

$$\int_0^x (x-t)^{\theta-1}\Delta_\theta(t|F_X)dt$$

2. when $\theta = 1$, the density is,

$$\xi_{F_X}(x) = \Delta_1(x|F_X) = \frac{1}{\pi}\sin(\pi F_X(x))\mathrm{e}^{-\Phi_{F_X}(x)}.$$

3. Density formulae for $\theta > 1$ are described as

$$\xi_{\theta F_X}(x) = (\theta-1)\int_0^x (x-t)^{\theta-2}\Delta_\theta(t|F_X)dt.$$

Result James, Lijoi and Prünster(2006)

Recognizing a connection of [CR]'s result for the cdf to Abel transforms, [JLP] reasoned that it is rather straightforward to establish the following

1. for all $\theta > 0$

$$\xi_{\theta F_X}(x) = \int_0^x (x-t)^{\theta-1} \Delta'(t|F_X)dt$$

2. Δ' is the derivative of Δ

In general the only really nice expression for the density is when $\theta = 1$

$$\xi_{F_X}(x) = \Delta_1(x|F_X) = \frac{1}{\pi}\sin(\pi F_X(x))\mathrm{e}^{-\Phi_{F_X}(x)}.$$

But notice that

$$\Delta_\theta(t|F_X) = \frac{1}{\pi}\sin(\pi\theta F_X(t))\mathrm{e}^{-\theta\Phi_{F_X}(t)} > 0.$$

for $0 < \theta \leq 1$ and all $t > 0$.

Comments on possible random variables that are Dirichlet means

1. The formulae we just saw are general expressions and don't look too familiar
2. However, we (James, Roynette and Yor-Coming Soon) show that there are many many familiar random variables that are Dirichlet means
3. This includes Generalized Inverse Gaussian, Positive Stable, Pareto, Uniform and at the very least a large class of random variables that are generalized gamma convolutions
4. Note if for instance U is uniform[0,1] then for $0 < \theta \leq 2$, there exist F_{X_θ} such that

$$U \stackrel{d}{=} M_\theta(F_{X_\theta})$$

A useful identity: Hjort and Ongaro 05

Let $D = (D_1, \ldots, D_k)$ denote a Dirichlet $(\theta_1, \ldots, \theta_k)$ random vector such that $\theta = \sum_{i=1}^{k} \theta_i$. Then

$$M_\theta(F_X) \stackrel{d}{=} \sum_{i=1}^{k} D_i M_{\theta_i}(F_X)$$

where $M_{\theta_i}(F_X)$ are independent and independent of D

Goal

Use the analytic work used to obtain results for $M_\theta(F_X)$ to obtain results for an entire family of Dirichlet mean random variables

$$\{M_\theta(F_{Z_c}) : c > 0\}$$

As a byproduct we get results for T_θ and Lévy processes derived from T_θ

Simple examples are of course the choices $Z_c = X + c$ and $Z_c = cX$, which, due to the linearity properties of mean functionals, results easily in the identities in law

$$M_\theta(F_{X+c}) = c + M_\theta(F_X) \text{ and } M_\theta(F_{cX}) = cM_\theta(F_X)$$

General DP Mean Identities

I introduced two simple ideas to address some of these points. The idea to construct such things was in part influenced quite a bit by some results of Pitman and Yor (1997) On length of excursions of bessel processes and others. Concerning multiplication by beta random variables and so on. The two results can be described as

1. Beta Scaling of DP Mean functionals
2. Gamma Tilting
3. Actually there are two more

Beta Scaling of DP mean functionals

Let $0 < \sigma \leq 1$ and $\theta > 0$. Let $\beta_{\theta\sigma,\theta(1-\sigma)}$ denote a beta random variable. Then consider the following simple random variable

$$\beta_{\theta\sigma,\theta(1-\sigma)} M_{\theta\sigma}(F_X)$$

Remember the density of $M_{\theta\sigma}(F_X)$ is not NICE except for $\theta\sigma = 1$.

Let Y_σ denote a Bernoulli random variable with success probability σ. Consider the independent product XY_σ with cdf

$$F_{XY_\sigma}(x) = \sigma F_X(x) + (1-\sigma)I(x \geq 0)$$

Result 1: change of total mass

Then

$$\beta_{\theta\sigma,\theta(1-\sigma)} M_{\theta\sigma}(F_X) \stackrel{d}{=} M_{\theta}(F_{XY_{\sigma}})$$

Why?

$$\psi_{F_{XY_\sigma}}(\lambda) = \int_0^\infty \log(1+\lambda w) F_{XY_\sigma}(dw) = \mathbb{E}[\log(1+\lambda XY_\sigma)].$$

but is equal to

$$\sigma\psi_{F_X}(\lambda) = \sigma \int_0^\infty \log(1+\lambda x) F_X(dx) = \sigma\mathbb{E}[\log(1+\lambda X)].$$

One to many

- For every fixed θ how many mean functionals of order θ did we just create?
- Answer: An uncountable number indexed by $0 < \sigma \leq 1$

$$(M_\theta(F_{XY_\sigma}) : 0 < \sigma \leq 1)$$

One to many explicit densities

- How many mean functionals of order $\theta = 1$ did we just create?
- Answer: An uncountable number indexed by $0 < \sigma \leq 1$

$$(M_1(F_{XY_\sigma}) : 0 < \sigma \leq 1)$$

$$\Phi_{F_{XY_\sigma}}(x) = \mathbb{E}[\log(|x - XY_\sigma|)I(XY_\sigma \neq x)] = \sigma\Phi_{F_X}(x) + (1-\sigma)\log(x)$$

Result 2

The density of

$$\beta_{\sigma,1-\sigma} M_\sigma(F_X) \stackrel{d}{=} M_1(F_{XY_\sigma})$$

is

$$\xi_{F_{XY_\sigma}}(x) = \frac{x^{\sigma-1}}{\pi} \sin(\pi F_{XY_\sigma}(x)) \mathrm{e}^{-\sigma \Phi_{F_X}(x)}$$

Note you do not need to know the density of $M_\sigma(F_X)$.

Back to GGC

Recall

$$T_{\theta\sigma} = G_{\theta\sigma} M_{\theta\sigma}(F_X)$$

but

$$G_{\theta\sigma} \stackrel{d}{=} G_{\theta}[\beta_{\theta\sigma,\theta(1-\sigma)}]$$

Hence

$$T_{\theta\sigma} = G_{\theta\sigma} M_{\theta\sigma}(F_X) = G_{\theta}[\beta_{\theta\sigma,\theta(1-\sigma)} M_{\theta\sigma}(F_X)]$$

Hence

$$T_{\theta\sigma} = G_{\theta\sigma} M_{\theta\sigma}(F_X) = G_\theta M_\theta(F_{XY_\sigma})$$

i.e. it is a GGC(θ, F_{XY_σ})

Hence

$$T_\sigma = G_\sigma M_\sigma(F_X) = G_1 M_1(F_{XY_\sigma})$$

i.e. it is a GGC$(1, F_{XY_\sigma})$

Exploiting infinite divisibility

Let

$$0 < \sum_{k=1}^{\infty} \sigma_k = 1$$

Then

$$T_1 \stackrel{d}{=} \sum_{k=1} T_{\sigma_k}$$

where T_{σ_k} are independent and each has distribution

$$T_{\sigma_k} \stackrel{d}{=} G_1 M_1(F_{XY_{\sigma_k}})$$

GGC species sampling models

Let (Z_k) denote iid random elements with some common distribution H. Then set

$$P_k = T_{\sigma_k}/T_1$$

Define a random probability measure as,

$$\sum_{k=1}^{\infty} P_k \delta_{Z_k}(dx) = P(dx)$$

THESE ARE NOT NRMI

That is, for any finite k

$$(P_1, \ldots, P_k) \stackrel{d}{=} (G_{1,i} M_1(F_{XY_{\sigma_i}})/[T' + S]; i = 1, \ldots, k)$$

where

$$T' = \sum_{i=1}^{k} G_{1,i} M_1(F_{XY_{\sigma_i}})$$

and for $\sigma_k^* = 1 - \sum_{i=1}^{k} \sigma_i$

$$S \stackrel{d}{=} G_1 M_1(F_{XY_{\sigma_k^*}})$$

independent of T'.

For fun one could choose for $i = 1, 2, \ldots$

$$\sigma_i = \lambda^{i-1} \mathrm{e}^{-\lambda} / (i-1)!$$

Fidi of GGC NRM

Let $\mu(\cdot)$ denote a completely random measure derived from T_1, such that

$$\mathbb{E}[\mu(A)] = H(A)$$

and $\mu(\mathcal{X}) = T_1$ Then for any A, set $H(A) = \sigma$, and one has that

$$\mu(A) = G_1 M_1(F_{XY_\sigma})$$

moreover for a disjoint partition of some space $\mathcal{X}$, $(C_1, \ldots, C_N)$ The fidi of $\mu(\cdot)$ is composed of independent components so that

$$\mu(C_i) = G_1 M_1(F_{XY_{\sigma_i}})$$

but writing

$$T_1 = \sum_{k=1}^{N} \mu(C_i)$$

This gives the fidi of the random probability measure,

$$P(\cdot) = \mu(\cdot)/T_1$$

i.e. $(P(C_1), \ldots, P(C_N).)$ These NRM have appeared in James Lijoi and Pruenster (2005) but not the fidi part.

Now conditioning on $T_1 = t$, this gives the fidi of the corresponding conditional Poisson Kingman class,

$$\mathcal{L}((P(C_1), \ldots, P(C_N)|T_1 = t)$$

Now mix over an arbitrary density $\gamma(t)$. Hence I just created an uncountable number of random probability measures with explicit fidi's.

Approximating GGC NRM's

Now suppose that for general θ, let $\mu_\theta(\cdot)$ denote a CRM derived from T_θ. Then form the NRM by,

$$P_\theta(\cdot) = \mu_\theta(\cdot)/T_\theta$$

Now write

$$T_\theta = \sum_{k=1}^{N} T_{\theta/N,k}$$

where $T_{\theta/N,k}$ are iid equal in distribution to $T_{\theta/N}$

Now we are interested in choosing $N > \theta$, in this case the joint distribution of

$$(T_{\theta/N,k}/T_\theta; k \leq N)$$

can be computed and certainly easily simulated. Since

$$T_{\theta/N} \stackrel{d}{=} G_1 M_1(F_{XY_{\theta/N}})$$

Furthermore as $N \to \infty$

$$\sum_{k=1}^{N} \frac{T_{\theta/N,k}}{T_\theta} \delta_{Z_k}(\cdot) \Rightarrow P_\theta(\cdot)$$

We can also use the left hand side as a sieve by letting N grow with the data.

Special Known Case

$G_{\theta/N,k}$ iid gamma$(\theta/N, 1)$ then

$$\sum_{k=1}^{N} \frac{G_{\theta/N,k}}{G_\theta} \delta_{Z_k}(\cdot) \Rightarrow D_\theta(\cdot)$$

where D_θ is a Dirichlet process