A Monte Carlo approach to a divergence minimization problem (work in progress)

IGAIA IV, June 12-17, 2016, Liblice

Michel Broniatowski

Université Pierre et Marie Curie, Paris, France

June 13, 2016

Contents

An inferential principle for minimization

A sequence of random elements X_n with values in a measurable space $(T, \mathcal{T})$ satisfies a Large Deviation Principle with rate Φ whenever, for all measurable set $\Omega \subset T$ it holds

$$\begin{aligned} -\Phi\left(int\left(\Omega\right)\right) &\leq \lim\inf_{n\to\infty} \varepsilon_n \log \Pr\left(X_n \in \Omega\right) \\ &\leq \lim\sup_{n\to\infty} \varepsilon_n \log \Pr\left(X_n \in \Omega\right) \leq -\Phi\left(cl\left(\Omega\right)\right) \end{aligned}$$

for some positive sequence ε_n where $int\left(\Omega\right)$ (resp. $cl\left(\Omega\right)$) denotes the interior (resp. the closure) of Ω in T and $\Phi(\Omega) := \inf\left\{\Phi(t); t \in \Omega\right\}$. The $\sigma-$field $\mathcal{T}$ is the Borel one defined by a given basis on T. For subsets Ω in T such that

$$\Phi\left(int\left(\Omega\right)\right) = \Phi\left(cl\left(\Omega\right)\right) \tag{1}$$

it follows by inclusion that

$$-\lim_{n\to\infty} \varepsilon_n \log \Pr\left(X_n \in \Omega\right) = \Phi\left(int\left(\Omega\right)\right) = \Phi\left(cl\left(\Omega\right)\right) = \inf_{t\in\Omega} \Phi(t) = \Phi(\Omega). \tag{2}$$

Assume that we are given such a family of random elements $X_1, X_2, ..$ together with a set $\Omega \subset T$ which satisfies (1). Suppose that we are interested in estimating $\Phi(\Omega)$. Then, whenever we are able to simulate a family of replicates $X_{n,1}, .., X_{n,K}$ such that $\Pr(X_n \in \Omega)$ can be approximated by the frequency of those $X_{n,i}$'s in Ω, say

$$f_{n,K}(\Omega) := \frac{1}{K} card\left(i : X_{n,i} \in \Omega\right) \tag{3}$$

a natural estimator of $\Phi(\Omega)$ writes

$$\Phi_{n,K}(\Omega) := -\varepsilon_n \log f_{n,K}(\Omega).$$

We have substituted the approximation of the variational problem $\Phi(\Omega) := \inf(\Phi(\omega), \omega \in \Omega)$ by a much simpler one, namely a Monte Carlo one, defined by (3).
No need to identify the set of points ω in Ω which minimize Φ.

This program can be realized whenever we can identify the sequence of random elements X_i's for which, given the criterion Φ and the set Ω, the limit statement (2) holds.

Here the X_i's are empirical measures of some kind, and $\Phi(\Omega)$ writes $\phi(\Omega, P)$ which is the infimum of a divergence between some reference probability measure P and a class of probability measures Ω.

Standpoint:

$$\phi(\Omega, P) \text{ is a LDP rate for specific } X_i\text{'s to be built.}$$

Applications: choice of models, estimation of the minimizers (dichotomy, etc)

Divergences

Let $(\mathcal{X}, \mathcal{B})$ be a measurable Polish space and P be a given reference probability measure (p.m.) on $(\mathcal{X}, \mathcal{B})$. Denote $\mathcal{M}_1$ the set of all p.m.'s on $(\mathcal{X}, \mathcal{B})$. Let φ be a proper closed convex function from $]-\infty, +\infty[$ to $[0, +\infty]$ with $\varphi(1) = 0$ and such that its domain $\mathrm{dom}\varphi := \{x \in \mathbb{R} \text{ such that } \varphi(x) < \infty\}$ is a finite or infinite interval . For any measure Q in $\mathcal{M}_1$, the ϕ-divergence between Q and P is defined by

$$\phi(Q, P) := \int_{\mathcal{X}} \varphi\left(\frac{dQ}{dP}(x)\right) \, dP(x).$$

if $Q << P$. When Q is not a.c. w.r.t. P, set $\phi(Q, P) = +\infty$. The ϕ-divergences between p.m.'s were introduced in Csiszar (1963) as "f-divergences" with some different definition.

For all p.m. P, the mappings $Q \in \mathcal{M} \mapsto \phi(Q, P)$ are convex and take nonnegative values. When $Q = P$ then $\phi(Q, P) = 0$. Furthermore, if the function $x \mapsto \varphi(x)$ is strictly convex on a neighborhood of $x = 1$, then

$$\phi(Q, P) = 0 \text{ if and only if } Q = P.$$

Cressie-Read divergences

When defined on $\mathcal{M}_1$, divergences associated with $\varphi_1(x) = x\log x - x + 1$ (KL), $\varphi_0(x) = -\log x + x - 1$ (KL_m-likelihood), $\varphi_2(x) = \frac{1}{2}(x-1)^2$ (Spearman Chi-square), $\varphi_{-1}(x) = \frac{1}{2}(x-1)^2/x$ (modified Chi-square, Neyman), $\varphi_{1/2}(x) = 2(\sqrt{x}-1)^2$ (Hellinger)
The class of Cressie and Read power divergences

$$x \in]0, +\infty[\mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)} \tag{4}$$

Extensions

The power divergences functions $Q \in \mathcal{M}_1 \mapsto \phi_\gamma(Q, P)$ can be defined on the whole vector space of signed finite measures $\mathcal{M}$ via the extension of the definition of the convex functions φ_γ : For all $\gamma \in \mathbb{R}$ such that the function $x \mapsto \varphi_\gamma(x)$ is not defined on $]-\infty, 0[$ or defined but not convex on whole $\mathbb{R}$, we extend its definition as follows

$$x \in]-\infty, +\infty[\mapsto \begin{cases} \varphi_\gamma(x) & \text{if} \quad x \in [0, +\infty[, \\ +\infty & \text{if} \quad x \in]-\infty, 0[. \end{cases} \tag{5}$$

Note that for the χ^2-divergence for instance, $\varphi_2(x) := \frac{1}{2}(x-1)^2$ is defined and convex on whole $\mathbb{R}$.

The conjugate (or Legendre transform) of φ will be denoted φ^*,

$$t \in \mathbb{R} \mapsto \varphi^*(t) := \sup_{x \in \mathbb{R}} \{tx - \varphi(x)\},$$

Property: φ is essentially smooth iff φ^* is strictly convex; then,

$$\varphi^*(t) = t{\varphi'}^{-1}(t) - \varphi\left({\varphi'}^{-1}(t)\right) \quad \text{and} \quad {\varphi^*}'(t) = {\varphi'}^{-1}(t).$$

In the present setting this holds.

The bootstrapped empirical measure

Let $Y, Y_1, Y_2, \ldots$denote a sequence of positive i.i.d. random variables . We assume that Y satisfies the so-called Cramer condition

$$\mathcal{N} := \left\{ t \in \mathbb{R} \text{ such that } \Lambda_Y(t) := \log Ee^{tY} < \infty \right\}$$

contains a neigborhood of 0 with non void interior.
Consider the weights $W_i^n, 1 \leq i \leq n$

$$W_i^n := \frac{Y_i}{\sum_{i=1}^n Y_i}$$

which define a vector of exchangeable variables $(W_1^n, .., W_n^n)$ for all $n \geq 1$.

The data $x_1^n, .., x_n^n$: We assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i^n} = P$$

a.s. and we define the bootstrapped empirical measure of $(x_1^n, .., x_n^n)$ by

$$P_n^W := \frac{1}{n} \sum_{i=1}^{n} W_i^n \delta_{x_i^n}.$$

A Sanov type result for the weighted Bootstrap empirical measure

Define the Legendre transform of Λ_Y, say Λ^* defined on $\operatorname{Im}\Lambda'$ by

$$\Lambda^*(x) := \sup_t tx - \Lambda_Y(t).$$

Theorem

Under the above hypotheses and notation the sequence P_n^W obeys a LDP on the space of all finite signed measures on X equipped with the weak convergence topology with rate function

$$\phi(Q,P) := \begin{cases} \inf_{m>0} \int \Lambda^*\left(m\frac{dQ}{dP}(x)\right) dP(x) & \text{if } Q << P \\ +\infty & \text{otherwise} \end{cases} \quad (6)$$

This Theorem is a variation on Corollary 3.3 in Trashorras and Wintenberger (2014).

Estimation of the minimum of the Kullback divergence

Set $Y_1, .., Y_n$ i.i.d. standard exponential . Then

$$\Lambda^*(x) = \varphi_1(x) := x \log x - x + 1$$

and

$$\inf_{m>0} \int \Lambda^* \left(m \frac{dQ}{dP}(x) \right) dP(x) = \int \Lambda^* \left(\frac{dQ}{dP}(x) \right) dP(x) = KL(Q, P).$$

Repete sampling $(Y_1, .., Y_n)$ i.i.d. E(1) K times. Hence for sets Ω such that

$$KL(int\Omega, P) = KL(cl\Omega, P)$$

then for large K

$$\frac{1}{n} \log \frac{1}{K} card \left\{ \left(P_n^W \right)_j \in \Omega, 1 \leq j \leq K \right\}$$

is a proxy of

$$\frac{1}{n} \log \Pr \left(P_n^W \in \Omega \right)$$

and therefore an estimator of $KL(\Omega, P)$.

When Y is E(1) then by Pyke's Theorem, $(W_1.., W_n)$ coincides with the spacings of the ordered statistics of n i.i.d. uniformly distributed r.v's on $(0,1)$, i.e. the simplest bootstrap version of P_n based on exchangeable weights.
It also holds with these weights

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\left(P_n^W \in \Omega \middle| x_1^n, .., x_n^n \right) - \frac{1}{n} \log \Pr\left(P_n \in \Omega \right) = 0$$

This weighted bootstrap is the only LDP efficient one.

Estimation of the minimum of the Likelihood divergence

$$KL_m(Q,P) := \int \varphi_0\left(\frac{dQ}{dP}\right) dP = -\int \log\left(\frac{dQ}{dP}\right) dP$$

$\varphi_0(x) := -\log x + x - 1.$
Set $Y_1, .., Y_n$ i.i.d. Poisson (1), then

$$\Lambda^*(x) = \varphi_0(x) := -\log x + x - 1$$

$$\inf_{m>0} \int \Lambda^*\left(m\frac{dQ}{dP}(x)\right) dP(x) = \int \Lambda^*\left(\frac{dQ}{dP}(x)\right) dP(x) = KL(Q,P).$$

Repete sampling $(Y_1, .., Y_n)$ i.i.d. Poisson(1) K times. For large K

$$\frac{1}{n}\log\frac{1}{K} card\left\{\left(P_n^W\right)_j \in \Omega, 1 \le j \le K\right\}$$

is an estimator of $KL_m(\Omega,P)$, since a proxy of

$$\frac{1}{n}\log \Pr\left(P_n^W \in \Omega\right)$$

A more general LDP associated with other divergences

We may also consider some wild bootstrap version, defining the wild empirical measure by

$$P_n^{Wild} := \frac{1}{n}\sum_{i=1}^{n} Y_i \delta_{x_i}$$

where the r.v's Y_1, $Y_{2,}$.. are i.i.d. with common expectation 1 and satisfy a Cramer condition with cumulant g.f.Λ_Y.

In this case it is somehow easy to prove the following general result.

Theorem

The wild empirical measure P_n^{Wild} obeys a LDP in the class of all signed finite measures endowed by the $\tau-$topology with good rate function $\phi\left(Q,P\right) = \int \Lambda^*\left(dQ/dP\right) dP$; Barbe and Bertail (1995); Najim (2005), etc

For adequate sets in the class of signed finite measures

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left(P_n^{Wild} \in \Omega \middle| x_1^n, .., x_n^n \right) = \phi\left(\Omega, P\right) \qquad (7)$$

Question

Is it possible to build r.v's $Y_1, .., Y_n$ such that

$$\lim_{n\to\infty} -\frac{1}{n} \log \Pr\left(P_n^{Wild} \in \Omega \middle| x_1^n, .., x_n^n \right) = \phi\left(\Omega, P\right)$$

holds for a given

$$\phi\left(Q, P\right) = \int \varphi\left(\frac{dQ}{dP}\right) dP$$

If yes then for "good" sets Ω , for large K

$$\frac{1}{n} \log \frac{1}{K} card \left\{ \left(P_n^{Wild} \right)_j \in \Omega, 1 \leq j \leq K \right\}$$

estimates $\phi\left(\Omega, P\right)$, since a proxy of

$$\frac{1}{n} \log \Pr\left(P_n^{Wild} \in \Omega \right)$$

Set of measures Ω to be considered in may satisfy

$$\phi(int\,(\Omega)\,, P) = \phi(cl\,(\Omega)\,, P) \tag{8}$$

where $int\,(\Omega)$ and $cl\,(\Omega)$ respectively denote the interior and the closure of the set Ω in $\mathcal{M}_1$ endowed with the corresponding τ or weak topology. Such sets Ω have been considered in the Large Deviation literature. Some sufficient conditions for (8) to hold; see Groeneboom, Osterhoof Ruymgaart (1979) (among others) for discussions. This is an entire field of questions and (counter) examples.
Estimation of $\phi(\Omega, P)$ is somehow an open problem: Usually try to identify the minimizers; difficult cases: Ω defined by moments of $L-$Statistics, ..Here find $\widehat{\phi(\Omega, P)}$ and get the minimizers after (dichotomy on Ω, etc).

A reciprocal statement to the LDP Theorem. We prove that any Cressie-Read divergence function is the Fenchel-Legendre transform of some moment generating function Λ. Henceforth we state a one to one correspondence between the class of Cressie-Read divergence functions and the distribution of some Y which can be used in order to build a bootstrap empirical measure of the form P_n^W.

We turn to some results on exponential families ; see Letac and Mora (1990).

Natural exponential families and their variance function

Given a positive measure μ on $\mathbb{R}$ consider the integral $\phi_\mu(t) := \int e^{tx} d\mu(x)$ and its domain $\mathcal{D}_\mu$, the set of all values of t such that $\phi_\mu(t)$ is finite, which is a convex (possibly void) subset of $\mathbb{R}$. Denote $k_\mu(t) := \log \phi_\mu(t)$ and let $m_\mu(t) := (d/dt)\, k_\mu(t)$ and $s_\mu^2(t) := \left(d^2/dt^2\right) k_\mu(t)$. Associated with μ is the Natural Exponential Family NEF(μ) of distributions

$$dP_t^\mu(x) := \frac{e^{tx} d\mu(x)}{\phi_\mu(t)}$$

which is indexed by t. It is a known fact that, denoting X_t a r.v. with distribution P_t^μ it holds $EX_t = m_\mu(t)$ and $VarX_t = s_\mu^2(t)$.
The NEF(μ) is *generated* by μ. NEF(ν) =NEF(μ) iff $d\nu(x) = \exp(ax + b)d\mu(x)$. This class is denoted NEF($\mathcal{B}$).

Defined on $\mathrm{Im}\, m_{\mathcal{B}}$ (all m_μ in $\mathcal{B}$ have same image), the function

$$x \to V(x) := s_\mu^2 o m_\mu^{\leftarrow}(x)$$

is independent of the peculiar choice of μ in $\mathcal{B}$ and is therefore called the *variance function* of the NEF($\mathcal{B}$).

Theorem

The function V characterizes the NEF, and reciprocally.

Starting with Morris (1982) a wide effort has been developed in order to characterize the basis of a NEF with given variance function. Stats: heteroscedastic models, variance regressed on the expectation; Tweedie (1947),...

Power variance functions

Power variance functions $V(x) = Cx^{\alpha}$ have been explored by various authors (BarLev and Ennis (1986), etc). NEF with variance function V are obtained through integration and identification of the resulting moment generating function. They are generated as follows (we identify the bases).

- For $\gamma < 0$ by stable distributions on $\mathbb{R}^+$ with characteristic exponent in $(0, 1)$. The resulting distributions define the Tweedie scale family (with base these stable laws) Example in the NEF: Inverse Gaussian $(\gamma = -1/2)$
- For $\gamma = 0$ by the exponential distribution
- For $0 < \gamma < 1$ by Compound Gamma-Poisson distributions
- For $\gamma = 1$ by the Poisson distribution
- For $\gamma = 2$ by the normal distribution

Other values of γ do not yield NEF's.

Theorem

(BarLev, Ennis) All distributions with power variance function are indefinitely divisible.

Consequence: a major tool for the simulation of the weights, etc*

Fact

The second derivative of the Legendre transform of the cumulant g.f. is the inverse of the variance function

Cressie-Read divergences , weights and variance functions

For

$$\varphi_\gamma(x) := C\frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$$

Any Cressie-Read divergence function is the Fenchel Legendre transform of a moment generating function of a random variable with expectation 1 and variance $1/C$ in a specific NEF, depending upon the divergence. Let Y be a r.v. with $\psi(t) := \log E \exp tY$ and power variance function

$$V(x) = \frac{1}{C}x^\alpha$$

. Then

$$\varphi_\gamma(x) = \psi^*(x) = \sup_t tx - \psi(t);$$

with $\alpha = 2 - \gamma$ .The NEF is generated by the distribution of Y. Since the differential equation $\frac{d^2}{dx^2}\varphi_\gamma(x) = Cx^{-\alpha}$ defines $\varphi_\gamma(x)$ in a unique way: one to one correspondence between Cressie-Read divergences and NEF's with power Variance functions. **Hence to any Cressie Read divergence its family of weights.**

Example

The Tweedie scale of distributions defines random variables Y with expectation 1 and variance C_τ corresponding to Cressie Read divergences with negative index $\gamma = -\tau/(1-\tau)$. The generator of the NEF (a measure μ) has characteristic function

$$f(t) = \exp\left\{iat - c\,|t|^{\tau}\,(1 + i\beta sign\,(t)\,\omega\,(t,\tau))\right\}$$

where $a \in \mathbb{R}$, $c > 0$ and $\omega(t,\tau) = \tan\left(\frac{\pi\tau}{2}\right)$ for $\tau \neq 1$, and $\omega(t,\tau) = \frac{2}{\pi}$ for $\tau = 1$.

Example

We consider the case when $\beta = 1$ and $0 < \tau < 1$ corresponding to a stable distribution on $\mathbb{R}^+$.For $\gamma = -1$,($\tau = 1/2$) the resulting divergence is

$$\varphi_{-1}(x) = \frac{1}{2}\frac{(x-1)^2}{x}$$

which is the modified χ^2 divergence (or Neyman χ^2). The associated r.v. Y has an Inverse Gaussian distribution with expectation 1 and variance 1.

Example

For $\gamma = 2$ it holds

$$\varphi_2(x) = \frac{1}{2}(x-1)^2$$

which is the Spearman χ^2 divergence. The resulting r.v. Y has a Gaussian distribution with expectation 1 and variance 1. Note that in this case, Y is not a positive random variable.

Example

For $\gamma = 1/2$ we get

$$\varphi_{1/2}(x) = 2\left(\sqrt{x} - 1\right)^2$$

which is the Hellinger divergence. The associated random variable Y has a Compound Gamma-Poisson distribution .

Example

When $\gamma = 3/2$ the distribution of Y belongs to the NEF generated by the stable law μ on $\mathbb{R}^+$ with characteristic exponent $1/3$,

$$f(x) = (d\mu(x)/dx) = (2\pi)^{-1}\lambda K_{1/2}\left(\lambda x^{1/2}\right)\exp\left(-px + 3\left(\lambda^2 p/4\right)^{1/3}\right)$$

where λ and p are positive and $K_{1/2}(z)$ is the modified Bessel function of order $1/2$ with argument z.

Example

When $\gamma = 1$ then

$$\varphi_0(x) = x \log x - x + 1,$$

the Kullback-Leibler divergence function, and Y has an exponential distribution with parameter 1.

Example

When $\gamma = 0$ then

$$\varphi_0(x) = -\log x + x - 1,$$

the Likelihood divergence and Y has a Poisson distribution with parameter 1.

Rare events, conditional limit results

$P_n^{Wild} \in \Omega$ may be a (very) rare event.
Consider

$$\frac{1}{K}\sum_{j=1}^{K} 1_\Omega\left(\left(P_n^{Wild}\right)_j\right)$$

Calculation may be long when $\Pr\left(\left(P_n^{Wild}\right) \in \Omega\right)$ is small (hit rate very low). This opens a range of questions.

Importance Sampling

Recall Let X some random element; assume it has a density p. We want to evaluate

$$\mathbf{P} := \Pr\left(X \in A\right)$$

Let $X_1, .., X_K$ be K independent copies of X and

$$P_K := \frac{1}{K}\sum_{i=1}^{K} 1_A\left(X_i\right)$$

the "naive" estimator of $\mathbf{P}$. For any density g where it makes sense

$$\mathbf{P} = \int 1_A\left(x\right) p(x) dx = \int 1_A\left(x\right) \frac{p(x)}{g(x)} g(x) dx$$

and therefore

$$P_{g,K} := \frac{1}{K}\sum_{i=1}^{K} 1_A\left(Z_i\right) \frac{p(Z_i)}{g(Z_i)}$$

converges to $\mathbf{P}$.

"MetaTheorem" The closer the sampling density to the density of X given $X \in A$, the most "efficient " the estimator. i.e. the highest the hit rate, the smallest the variance, etc.

Here

$$X := P_n^{Wild} = \frac{1}{n} \sum_{i=1}^{n} Y_i \delta_{x_i^n}.$$

Assume for example that

$$\Omega := \left\{ Q : \int f(x) dQ(x) > s \right\}$$

say for some f and real a.

$$\left(P_n^{Wild} \in \Omega \right) = \left(\frac{1}{n} \sum_{i=1}^{n} Y_i f(x_i^n) > s \right).$$

With $x_i^n = x_i$ and $f(x_i) = a_i$

$$\left(P_n^{Wild} \in \Omega \right) = \left(\frac{1}{n} \sum_{i=1}^{n} a_i Y_i > s \right)$$

The form of the estimator $(Y_{1,1},..,Y_{n,1}),...,(Y_{n,1},..,Y_{n,K})$ i.i.d samples of i.i.d. replications

$$P_K := \frac{1}{K}\sum_{i=1}^{K} 1_{(s,\infty)}\left(\frac{1}{n}\sum_{i=j}^{n} a_j Y_{j,i}\right)$$

The IS estimator

$$P_{g,K} := \frac{1}{K}\sum_{i=1}^{K} 1_{(s,\infty)}\left(\frac{1}{n}\sum_{j=1}^{n} a_j Z_{j,i}\right)\frac{p(Z_{1,i})...p(Z_{n,i})}{g(Z_{1,i},..,Z_{n,i})}$$

where g is any density on $\mathbb{R}^n$ where the ratio is defined.

Approximate the density of $(Y_1, .., Y_n)$ given $\left(\frac{1}{n}\sum_{j=1}^{n} a_j Y_i > s\right)$; Gibbs conditional result (Csiszar, 1984), Dembo, Zeitouni (1996), Br-Caron (2014),etc.

Exploring the minimizers of $\phi(Q, P)$ when $Q \in \Omega$ and $\Omega := \cup_\alpha \Omega_\alpha, \alpha \in A$

Dichotomy:

Estimate $\phi(\Omega, P)$.

Split A into A_1 and A_2 so that $\Omega = \Omega^1 + \Omega^2$

$(\Omega^j := Q \in \Omega$: there exists some α in A_j with $Q \in \Omega_\alpha$.

Estimate $\phi\left(\Omega^1, P\right)$ and $\phi\left(\Omega^2, P\right)$

If $\phi(\Omega, P) = \phi\left(\Omega^j, P\right)$ then a minimizer is in Ω^j.

Split these Ω^j and iterate

Example

The Tweedie scale

Let Z be a r.v. with stable distribution on $\mathbb{R}^+$ and density p . Its characteristic function $f(t) = E \exp itZ$ is described by the formula

$$f(t) = \exp\left\{iat - c\,|t|^{\tau}\,(1 + i\beta sign(t)\,\omega(t,\tau))\right\}$$

where $a \in \mathbb{R}$, $c > 0$ and $\omega(t,\tau) = \tan\left(\frac{\pi\tau}{2}\right)$ for $\tau \neq 1$, and $\omega(t,\tau) = \frac{2}{\pi}$ for $\tau = 1$.

We consider the case when $\beta = 1$ and $0 < \tau < 1$ corresponding to a stable distribution on $\mathbb{R}^+$ which therefore satisfies the following characterization: For $Z_1, ..., Z_n$ n i.i.d. copies of Z there exists $a_n > 0$ such that

$$\frac{Z_1 + ... + Z_n}{a_n} =_d Z$$

where the equality holds in distribution. Also $a_n = n^{1/\tau}$. The Laplace transform of p satisfies

$$\varphi(t) := \int_0^\infty e^{-tx} p(x) dx = e^{-t^\tau}$$

for all non negative value of t; see [**?**].

Associated with p is the Natural Exponential family (NEF) with basis p namely the densities defined for non negative t through

$$p_t(x) := e^{-tx} p(x) / e^{-t^\tau}$$

with support $\mathbb{R}^+$. For positive t , a r.v. X_t with density p_t has a moment generating function $E \exp \lambda X_t$ which is finite in a non void neighborhood of 0 and therefore has moments of any order.

Consider the density $p_1(x) = e^{-x+1} p(x)$ with finite m.g.f. in $(-\infty, 1)$, expectation $\mu = \tau$ and variance $\sigma^2 = \tau(1-\tau)$. Finally set for all non negative x

$$q(x) := \sqrt{\tau(1-\tau)} p_1 \left(x \sqrt{\tau(1-\tau)} + \tau - 1 \right)$$

which is for all $0 < \tau < 1$ the density of some r.v. Y with expectation 1 and variance 1. The m.g.f. of Y is

$$E \exp \lambda Y = e \exp \left[1 - \frac{\tau}{\sqrt{\tau(1-\tau)}} \right] \exp - \left[1 - \frac{\lambda}{\sqrt{\tau(1-\tau)}} \right]^\tau .$$

For $\tau = 1/2$, Y has the Inverse Gaussian distribution with parameters $(1, 1)$ and m.g.f

$$E \exp \lambda Y = e \left(\exp - [1 - 2\lambda]^{1/2} \right).$$

The variance function of the NEF generated by a stable distribution with index τ in $(0, 1)$ writes

$$V(x) = C_\tau x^{\frac{2-\tau}{1-\tau}}$$

with

$$C_\tau := \left(\frac{1 - \tau}{\tau} \right)^{\frac{2-\tau}{2(1-\tau)}} .$$

Example

Compound Gamma Poisson distributions

We briefly characterize this compound distribution and the resulting weight W. Let μ denote the distribution of $S_N := \sum_{i=0}^{N} \Gamma_i$ where $S_0 := 0$, N is a Poisson (p) r.v. independent of the independent family $(\Gamma_i)_{i>1}$

where the Γ_i's are distributed with Gamma distribution with scale parameter $1/\lambda$ and shape parameter $-\rho$. Here

$$\rho := \frac{\gamma - 1}{\gamma}$$

$$\lambda := \rho$$

$$p := (\gamma - 1)^{-1/\gamma}$$

where we used the results in [?] p1516. Consider the family of distributions NEF(μ) generated by μ, which has power variance function $V(x) = x^{\gamma+1}$ defined on $\mathbb{R}^+$. The r.v. W has distribution in NEF(μ) with expectation and variance 1. Its density is of the form

$$f_W(x) := \exp(ax + b) f(x)$$

where $f(x) := (d\mu(x)/dx)$ is the density of S_N. The values of the parameters a and b are

$$a := -1$$

$$b := -(\gamma - 1)^{-1/\gamma} \left[\left(1 - \frac{\gamma}{\gamma - 1} \right)^{\rho} - 1 \right]$$