Generalized Likelihood Functions and Random Measures

Christos E. Kountzakis

Department of Mathematics
University of the Aegean
Karlovassi GR-83 200, Samos, Greece

Copyright © 2014 Christos E. Kountzakis. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The density estimators we construct in this paper, called *generalized likelihoods*, rely on specific random measures. The key idea is to approximate an unknown density function -given a sample of its observations through another function which is made by another family of known density functions. Under this point of view, we conclude that under some assumptions the classes of distributions, which constitute the maximum domains of attraction can be viewed to be two instead of three and we discuss the question of difference between the classes of equivalence of probability distributions.

Mathematics Classification: 62K05; 60GG57; 62K20

Keywords: Random measures; Density estimation

1 Introduction

Many methods for the estimation of an unknown density function f by means of functions of i.i.d random variables $X_1, X_2, \dots, X_n$ have been proposed in recent years. These include the kernel methods studied in [4], the characteristic function approach studied in [5] and the properties for Lévy Random measure studied by [3] have shown that in certain cases these estimates are best possible.

In [4], the problem of estimation of a probability density function was examined and the problem of determining the mode of a probability density

function, as well. The construction of a family of estimates of $f(x)$ was indicated, and of the mode, which are consistent and asymptotically normal. The problem of estimating the mode of a probability density function is somewhat similar to the problem of maximum likelihood estimation of a parameter.

Also, a random measure is said to be a submeasure of a second random measure if its probability law is absolutely continuous with respect to that of the second. In [3], it was proved that if the second measure is a Lévy random measure then the submeasure is Lévy if and only if the Radon-Nikodym derivative satisfies a natural factorization condition.

In this paper, we actually show that the Weibull and Fréchet distribution families do and do not belong respectively, to the equivalence class of densities defined by the kernel of Gumbel distributions. For this we use a random measure in accordance with the definition given in [3].

These families are important, since they are the Maximum Domains of Attraction, but we also discuss other distribution families, too.

2 Random Measures and Kernel Densities

We consider a probability space $(\Omega, \mathcal{F}, \mu)$ and a measurable space $(E, \mathcal{E})$. A **random measure** in [3] ξ on $(E, \mathcal{E})$ over the probability space $(\Omega, \mathcal{F}, \mu)$ is a map

$$\xi : \mathcal{E} \times \Omega \rightarrow \mathbf{R}_+,$$

such that:

1. the map $\omega \mapsto \xi(A, \omega)$ is a random variable for any $A \in \mathcal{E}$
2. the map $A \mapsto \xi(A, \omega)$ is a measure on $\mathcal{E}$, μ - almost surely in Ω .

In the next, the measurable spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ are identified to $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$, while $\mu = \lambda$, where λ denotes the Lebesgue measure on $\mathbf{R}$.

In [5] and [4] a function $K : \mathbf{R} \rightarrow \mathbf{R}$ is called **kernel** if it satisfies the following properties - in the next called **kernel conditions**:

$$(i) \sup\{|K(x)|, x \in \mathbf{R}\} < \infty, \int_{\mathbf{R}} |K(x)| dx < \infty,$$

$$(ii) \lim_{x \rightarrow \infty} |xK(x)| = 0, \int_{\mathbf{R}} K(x) dx = 1,$$

$$(iii) \int_{\mathbf{R}} x^i K(x) dx = 0, i = 1, 2, \dots, m-1, \int_{\mathbf{R}} |x|^{m-\frac{1}{p}} |K(x)| dx < \infty,$$

$$m \in \mathbf{N}, p \in \mathbf{R}_{++}.$$

The **Sobolev space** $W_p^{(m)}(M)$, where $M > 0, p > 0$ and $m \in \mathbf{N}$ is a function space over the domain $\mathbf{R}$ for example which consists of the functions

$$\{f : \mathbf{R} \rightarrow \mathbf{R} | f^{(k)},$$

is absolutely continuous for

$$k = 1, 2, \dots, m-1, f^{(m)} \in L^p(\mathbf{R}), \|f^{(m)}\|_p \leq M\}.$$

A function $f : I \rightarrow \mathbf{R}$, where I is an interval of $\mathbf{R}$ is absolutely continuous on I if for any $\varepsilon > 0$, there is a $\delta > 0$ such that whenever a finite family of sub-intervals $(x_k, y_k), k = 1, 2, \dots, m$ of I exists such that $\sum_{i=1}^m (y_i - x_i) < \delta$, then $\sum_{i=1}^m (f(y_i) - f(x_i)) < \varepsilon$ holds.

A function f is absolutely continuous on $\mathbf{R}$ if it is absolutely continuous on every interval of $\mathbf{R}$.

A criterion for the absolute continuity is the existence of the first derivative f' λ -almost everywhere and that f' is Lebesgue integrable.

The estimator of the density $f \in W_p^{(m)}(M)$ which relies on the ordered sample $t_1, t_2, \dots, t_n$ of it and also on the kernel K is equal to

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - t_i}{h_n}\right).$$

$h_n > 0$ is selected so that $h_n \rightarrow 0, nh_n \rightarrow \infty$. For example $h_n = \frac{1}{\sqrt{n}}, n \in \mathbf{N}$.

If we suppose that $m = 2, p = 1, K(x) \geq 0, x \in \mathbf{R}$, then the above conditions become

$$\sup\{K(x), x \in \mathbf{R}\} < \infty,$$

$$\lim_{x \rightarrow \infty} |x|K(x) = 0, \int_{\mathbf{R}} K(x)dx = 1,$$

$$\int_{\mathbf{R}} xK(x)dx = 0.$$

According to [5, Theorem 4.1],

$$\int_{\mathbf{R}} (f(x) - \hat{f}_n(x))^2 dx \rightarrow 0,$$

while $n \rightarrow \infty$ and for $m = 2, p = 1$ the above integral is less than $\frac{c}{n^{\frac{2}{3}}}$, where $c > 0$, if $f \in W_1^{(2)}$.

We may use other **density functions** which satisfy the previous conditions of the kernels in order to approximate another unknown density function, by taking observations of it.

We consider a random measure

$$\xi : \mathcal{B}_{\mathbf{R}} \times \mathcal{Y} \rightarrow \mathbf{R}_+,$$

such that:

1. the map $y \mapsto \xi(A, y)$ is a random variable for any $A \in \mathcal{B}_{\mathbf{R}}$
2. the map $A \mapsto \xi(A, y)$ is a measure on $\mathcal{E}$, λ - almost surely in $\mathcal{Y}$,
3. λ is the Lebesgue measure restricted on the open set $\mathcal{Y} \subseteq \mathbf{R}$.

Hence by this way for a single-parametric family of densities satisfying the **kernel conditions**, whose parametric space is the set $\mathcal{Y}$, we may define the random measure ξ_K whose the measure -part is

$$\xi_K(y)(x) = K_y(x), x \in \mathbf{R}, y \in \mathcal{Y}.$$

If we take a single-parametric family of density kernels whose parameter takes the value y or else if we take any measure-value of the random measure ξ_K on $\mathcal{Y}$, then the convergence result [5, Theorem 4.1], takes the form:

Theorem 2.1 *If $f(x)$ is an unknown density in $W_1^{(2)}$ and $t_1, t_2, \dots, t_n$ is an ordered sample from it, then*

$$\int_{\mathbf{R}} (f(x) - \hat{f}_n(x))^2 dx \rightarrow 0,$$

and if $m = 2, p = 1$ the above integral is less than $\frac{c_1}{n^{\frac{2}{3}}}$, where $c_1 > 0$ and

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K_y\left(\frac{x - t_i}{h_n}\right).$$

$h_n > 0$ is selected so that $h_n \rightarrow 0$, $nh_n \rightarrow \infty$ if $y \in \mathcal{Y}$ and every K_y satisfies the **kernel conditions**.

Proof: The proof arises from [5, Theorem 4.1].

Definition 2.2

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K_y\left(\frac{x - t_i}{h_n}\right)$$

is called **generalized likelihood function** for $f(x)$.

3 Classes of Equivalence for Densities

The closure of the class $K_{\mathcal{Y}} = \{K_y, y \in \mathcal{Y}, \mathcal{Y} \subseteq \mathbf{R}\}$, is

$$\mathcal{F}_K = \{f(x) \in W_{1,+}^{(2)} \mid \int_{-\infty}^{\infty} (f(x) - \hat{f}_{K_y,n})^2 \rightarrow 0, n \rightarrow +\infty\},$$

for some ordered sample $t_1, t_2, \dots, t_n, \dots$ from $f(x)$ and for some sequence $(h_n)_{n \in \mathbf{N}}$, such that $h_n > 0$ and $h_n \rightarrow 0$, while $nh_n \rightarrow +\infty$.

We may define a binary relation $\approx$ between the densities in $W_1^{(2)}$ according to the families of kernels, which satisfy the **kernel conditions** and approximate them. This relation is defined as follows: $f \approx g \Leftrightarrow f, g \in \mathcal{F}_K$.

Theorem 3.1 *The relation $\approx$ is an equivalence relation, namely reflexive, symmetric and transitive.*

Proof:

- (i) If $f \in \mathcal{F}_K$, then $f, f \in \mathcal{F}_K$, which implies $f \approx f$ (Reflexivity).
- (ii) If $f, g \in \mathcal{F}_K$, namely $f \approx g$, we notice that $g, f \in \mathcal{F}_K$, hence $g \approx f$ (Symmetry).
- (iii) If $f, g \in \mathcal{F}_K$, namely $f \approx g$, while $g, z \in \mathcal{F}_K$, namely $g \approx z$, we deduce that $f, z \in \mathcal{F}_K$, namely $f \approx z$.

Theorem 3.2 *The relation $\approx$ divides the set of the densities for which $f \in W_1^{(2)}$ into disjoint classes of equivalence.*

Proof: The proof of the Theorem relies on the fact that if an unknown density $f \in W_1^{(2)}$ belongs to a pair of different classes $\mathcal{F}_K, \mathcal{F}_L$, for the same $\varepsilon > 0$ we obtain:

$$\int_{\mathbf{R}} \left(\frac{1}{nh_n} \sum_{i=1}^n K_y \left(\frac{x - t_i}{h_n} \right) - f(x) \right)^2 dx < \frac{\varepsilon^2}{8}, n \geq n_1(\varepsilon),$$

$$\int_{\mathbf{R}} \left(\frac{1}{n\eta_n} \sum_{i=1}^n L_y \left(\frac{x - \tau_i}{\eta_n} \right) - f(x) \right)^2 dx < \frac{\varepsilon^2}{8}, n \geq n_2(\varepsilon),$$

where L is another kernel, $\tau_i, i = 1, 2, \dots, n$ are sample points from the unknown density, too $(\eta_n)_{n \in \mathbf{N}}$ is a sequence of real numbers such that $\eta_n > 0$ and $\eta_n \rightarrow 0$, while $n\eta_n \rightarrow +\infty$. We consider $n_0(\varepsilon) = \max\{n_1(\varepsilon), n_2(\varepsilon)\}$. For this $n_0(\varepsilon)$ and by the Minkowski Inequality, we obtain

$$\int_{\mathbf{R}} \left(\frac{1}{nh_n} \sum_{i=1}^n K_y \left(\frac{x - t_i}{h_n} \right) - \frac{1}{n\eta_n} \sum_{i=1}^n L_y \left(\frac{x - \tau_i}{\eta_n} \right) \right)^2 dx < \frac{\varepsilon^2}{4}, n = n_0(\varepsilon).$$

The last inequality implies that

$$\frac{1}{nh_n} \sum_{i=1}^n K_y\left(\frac{x-t_i}{h_n}\right) = \frac{1}{n\eta_n} \sum_{i=1}^n L_y\left(\frac{x-\tau_i}{\eta_n}\right), n = n_0(\varepsilon), x \in \mathbf{R},$$

a contradiction.

Theorem 3.3 *The family of normalized Gumbel distributions can be used as kernels for the approximation of a Weibull distribution according to Theorem 1.*

Proof: This arises from the fact that a normalized Gumbel density f ($\int_{\mathbf{R}} xf(x)dx = 0$) satisfies the kernel conditions and the abstract Weibull density $h \in W_1^{(2)}$. The fact that the normalized Gumbel densities $K(x) = \frac{1}{b} \exp(-z - \exp(-z))$, $z = \frac{x-m}{b}$, $b > 0$, $m \in \mathbf{R}$, $x \in \mathbf{R}$, being actually the ones with $\int_{\mathbf{R}} xK(x)dx = 0$, $\int_{\mathbf{R}} x^2K(x)dx = 1$, satisfy the rest kernel conditions $\sup\{K(x), x \in \mathbf{R}\} < \infty$, $\lim_{x \rightarrow \infty} |x|K(x) = 0$, $\int_{\mathbf{R}} K(x)dx = 1$, is verified as follows.: The third condition is verified by the fact that $K(x)$ is a density function. The first property is verified by the functional form of $K(x)$ and specifically from the inequality $|K(x)| \leq \exp(-\exp(-z))$. The second one is verified by the inequality $|x|K(x) \leq e^{z\frac{1}{b}} e^{-z} \exp(-\exp(-z)) = \frac{1}{b} \exp(-\exp(-z))$.

Corollary 3.4 *There is at least one non-heavy tailed distribution density approximated by heavy-tailed distributions in the sense of Theorem 1.*

Proof: Normalized Gumbel distributions belong to the class of heavy-tailed distributions. On the other hand, the Weibull distribution density $f(x) = \frac{k}{a} (\frac{x}{a})^{k-1} \exp(-(\frac{x}{a})^k)$, $x \geq 0$, for $k = 1$ is equal to an exponential distribution density, which is the density of a non-heavy tailed distribution.

About classes of heavy-tailed distributions, see for example in [2].

Theorem 3.5 *The family of Fréchet distributions cannot be approximated by normalized Gumbel distributions according to Theorem 1. This holds because if $k(x)$ is a Fréchet density, this does not belong to the Sobolev space $W_1^{(2)}$.*

Proof: The Fréchet distribution density is: $f(x) = \frac{a}{s} (\frac{x-m}{s})^{-a-1} \exp(-(\frac{x-m}{s})^{-a})$, $x > m$, $s > 0$. The first derivation of $f(x)$ is:

$$f'(x) = \left(\frac{a(\frac{-m+x}{s})^{-a-1}}{\exp(\frac{m-x}{s})^a} \right)' = -\frac{a(am+sa-ax+s)\exp(\frac{m-x}{s})^{-a} (\frac{x-m}{s})^{-a}}{s(m-x)^2}$$

$$f'(x) = -\frac{a(am+sa-ax+s)\exp(\frac{m-x}{s})^{-a} (\frac{x-m}{s})^{-a}}{s(m-x)^2}.$$

The second derivation of $f(x)$ is:

$$f''(x) = -\frac{a(a^2(m+s-x)^2 + as(2m+3s-2x) + 2s^2)\exp(\frac{m-x}{s})^{-a}(\frac{x-m}{s})^{-a}}{s^2(m-x)^3}.$$

For $x > m$ and $s > 0$:

$$\int_m^\infty -\frac{a(a^2(m+s-x)^2 + as(2m+3s-2x) + 2s^2)\exp(\frac{m-x}{s})^{-a}(\frac{x-m}{s})^{-a}}{s^2(m-x)^3}dx = \infty.$$

The Fréchet distribution density is not absolutely continuous on $\mathbf{R}$. So Fréchet density does not belong to the Sobolev space $W_1^{(2)}$.

Definition 3.6 If two distribution families K, L , are called **different** if there is at least one member(density) $f \in L$, such that $f \notin \mathcal{F}_K$ holds.

Corollary 3.7 (Normalized) Gumbel family is different from Fréchet family.

Proof: See the Proof of the previous Theorem.

We may provide some more Examples of difference of distribution families.

Theorem 3.8 Exponential family is different from the Gaussian family of distributions.

Proof: Let us consider the family of canonical kernels

$$K_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}x^2\},$$

where $\sigma > 0$. The kernel conditions

$$\sup\{K_\sigma(x), x \in \mathbf{R}\} < \infty, \lim_{x \rightarrow \infty} |x|K_\sigma(x) = 0,$$

$$\int_{\mathbf{R}} K_\sigma(x)dx = 1, \int_{\mathbf{R}} xK_\sigma(x)dx = 0$$

hold for the canonical kernels, hence we may examine if the exponential distribution $f(x) = u \cdot \exp\{-ux\}H_0(x)$, where $H_0(x)$ is the Heaviside step function, relative to 0 and $u > 0$. We notice that $f \in W_1^{(2)}$. We consider the ordered sample points $t_1, t_2, \dots, t_n$ from the exponential distribution, where

$$t_1 = \min\{t_i | i = 1, 2, \dots, n\} > 0.$$

We are going to show that the following non-convergence result is true:

$$\int_{\mathbf{R}} (\frac{1}{nh_n} \sum_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x-t_i)^2}{2\sigma^2h_n^2}\} - f(x))^2 dx \rightarrow +\infty, h_n = \frac{1}{\sqrt{n}}, n \rightarrow +\infty.$$

The above integral is greater or equal than the following sum of integrals:

$$\sum_{i=1}^n \left(\int_0^\infty \left(\frac{1}{2n\pi\sigma^2} \exp\left\{-\frac{nx^2}{\sigma^2}\right\} dx \right) = \sum_{i=1}^n \frac{1}{2n\pi\sigma^2} \frac{\sigma\sqrt{\pi}}{2\sqrt{n}} = \sum_{i=1}^n \frac{\sigma\sqrt{\pi}}{4n^{\frac{3}{2}}} = \infty. \right.$$

Hence, the family of exponential distributions is different than the one of the canonical ones.

We notice that both the canonical and the exponential distributions are not heavy-tailed.

Theorem 3.9 *The family of Pareto -type distributions is different from the Gaussian family of distributions.*

Let us consider the family of canonical kernels

$$K_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}x^2\right\},$$

where $\sigma > 0$. The kernel conditions

$$\sup\{K_\sigma(x), x \in \mathbf{R}\} < \infty, \lim_{x \rightarrow \infty} |x|K_\sigma(x) = 0,$$

$$\int_{\mathbf{R}} K_\sigma(x) dx = 1, \int_{\mathbf{R}} xK_\sigma(x) dx = 0$$

hold for the canonical kernels, hence we may examine if the Pareto-type distribution $g(x) = \frac{2}{x^3}H_1(x)$, where $H_1(x)$ is the Heaviside step function, relative to 1. We notice that $g \in W_1^{(2)}$. We consider the ordered sample points $t_1, t_2, \dots, t_n$ from the exponential distribution, where

$$t_1 = \min\{t_i | i = 1, 2, \dots, n\} > 0.$$

We are going to show that the following non-convergence result is true:

$$\int_{\mathbf{R}} \left(\frac{1}{nh_n} \sum_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-t_i)^2}{2\sigma^2h_n^2}\right\} - g(x) \right)^2 dx \rightarrow +\infty, \quad h_n = \frac{1}{\sqrt{n}}, \quad n \rightarrow +\infty.$$

The above integral is greater or equal than the following sum of integrals:

$$\sum_{i=1}^n \left(\int_1^\infty \left(\frac{4}{\sigma\sqrt{2n\pi}x^3 \exp\left\{\frac{nt_ix^2}{\sigma^2}\right\}} dx \right) \geq \sum_{i=1}^n \frac{2}{\sigma\sqrt{2n\pi} \exp\left\{\frac{1}{\sigma^2}\right\}} = +\infty.$$

We also notice that this Pareto -type distribution is a heavy-tailed one, while a Gaussian distribution is not a heavy-tailed distribution.

References

- [1] Aliprantis C.D., Border K.C. (1999). Infinite Dimensional Analysis, A Hitchhiker's Guide, (second edition). *Springer*.
- [2] Cai, J., Tang, Q. (2004). On max-sum equivalence and convolution closure of heavy-tailed distributions and their applications. *Journal of Applied Probability*, **41** 117-130.
- [3] Karr A. (1978). Lévy Random Measures, *Annals of Probability* **6**, 57–71.
- [4] Parzen E. (1962). On Estimation of a Probability Density and Mode. *Annals of Mathematical Statistics* **33**, 1065–1076.
- [5] Wahba G. (1975). Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation. *Annals of Statistics* **3**, 15–29.

Received: March 18, 2014