Electron. Commun. Probab. **21** (2016), no. 6, 1–8.
DOI: 10.1214/16-ECP4458
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

When does the minimum of a sample of an exponential family belong to an exponential family?

Shaul K. Bar-Lev* Gérard Letac†

Abstract

It is well known that if $(X_1, ..., X_n)$ are i.i.d. r.v.'s taken from either the exponential distribution or the geometric one, then the distribution of $\min(X_1, ..., X_n)$ is, with a change of parameter, is also exponential or geometric, respectively. In this note we prove the following result. Let F be a natural exponential family (NEF) on $\mathbb{R}$ generated by an arbitrary positive Radon measure μ (not necessarily confined to the Lebesgue or counting measures on $\mathbb{R}$). Consider n i.i.d. r.v.'s $(X_1, ..., X_n), n \geq 2$, taken from F and let $Y = \min(X_1, ..., X_n)$. We prove that the family G of distributions induced by Y constitutes an NEF if and only if, up to an affine transformation, F is the family of either the exponential distributions or the geometric distributions. The proof of such a result is rather intricate and probabilistic in nature.

Keywords: exponential family; exponential distribution; geometric distribution; order statistics; Radon measure.
AMS MSC 2010: 62E10.
Submitted to ECP on 2015-08-03, final version accepted on 2015-12-24.

1 Introduction

Both distributions, the geometric distribution supported on $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and the exponential distribution supported on $[0, \infty)$, possess similar properties. We outline only some of them:

- Like its continuous analogue (the exponential distribution), the geometric distribution is memoryless.
- If a r.v. X has an exponential distribution with mean $1/\lambda$ then $\lfloor X \rfloor$, where $\lfloor x \rfloor$ denotes the floor function of a real number x, is geometrically distributed with parameter $p = 1 - e^{-\lambda}$.
- If $(X_1, ..., X_n)$ are i.i.d. r.v.'s taken from either the exponential distribution or the geometric one, then the distribution of $\min(X_1, ..., X_n)$ is, with a change of parameter, also exponential or geometric, respectively.
- Both families of distributions belong to the class of natural exponential families (NEF's).

Indeed, the present note incorporates the last two properties in the following sense. Let F be an NEF on $\mathbb{R}$ generated by an arbitrary positive Radon measure μ (not necessarily confined to the Lebesgue or counting measures on $\mathbb{R}$). Consider n i.i.d. r.v.'s

*University of Haifa, Israel. E-mail: barlev@stat.haifa.ac.il
†Université Paul Sabatier, France. E-mail: gerard.letac@math.univ-toulouse.fr

$(X_1, ..., X_n), n \geq 2$, taken from F and let $Y = \min(X_1, ..., X_n)$. Then we prove that the family G of distributions induced by Y constitutes an NEF if and only if, up to an affine transformation, F is the family of either the exponential distributions or the geometric distributions.

A similar, but rather more restrictive, problem has been treated by Bar-Lev and Bshouty (2008) in which they considered the case where μ has the form $\mu(dx) = h(x)dx$. Then under some restrictive conditions on h (as differentiability) they showed that the family of distributions induced by Y is an NEF if and only if the distribution of the X_i's is an exponential one (up to an affinity $x \mapsto ax + b$). In their concluding remarks, Bar-Lev and Bshouty (2008) indicated the mathematical difficulties arising for proving that when μ is a counting measure on $\mathbb{N}_0$ then the family G is an NEF iff F is the family geometric distributions. It should be noted, however, that for the restricted case $\mu(dx) = h(x)dx$, Bar-Lev and Bshouty (2008) treated the question of when G_r, the family of distributions induced by the r-th order statistic $X_{(r)}$ (out $(X_1, ..., X_n)$), is an NEF. They showed that necessarily $r = 1$ in which case the NEF F must be that of the exponential distributions.

As already indicated, we consider here the case $r = 1$ and prove in Theorem 1 a more general result for an arbitrary measure μ (which includes the Lebesgue measure and counting measure as special cases).

In Section 2 we introduce some required preliminaries on NEF's. In Section 3 we present and prove our main result. The style of the result and the methods of the proof are close to the celebrated Balkema-de Haan-Pickands theorem on extreme values (see [1] and [5]).

2 Some preliminaries on NEF's

For proving our main result we shall need the definition of an NEF (for a detailed description of NEF's on $\mathbb{R}$ see Letac and Mora, 1990).

Let μ be a positive non-Dirac Radon measure on $\mathbb{R}$. The Laplace transform of μ is

$$L_\mu(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \mu(dx) \leq \infty.$$

Let

$$D(\mu) = \{\theta \in \mathbb{R} : L_\mu(\theta) < \infty\}, \quad \Theta(\mu) = \operatorname{int} D(\mu)$$

and denote $k_\mu(\theta) = \log L_\mu(\theta), \theta \in \Theta(\mu)$. Also, let $\mathcal{M}(\mathbb{R})$ denote the set of positive measures μ on $\mathbb{R}$ not concentrated on one point such that $\Theta(\mu) \neq \emptyset$. Then, the family of probabilities

$$F = F(\mu) = \{P(\theta, \mu) : \theta \in \Theta(\mu)\}$$

where

$$P(\theta, \mu)(dx) = e^{\theta x - k_\mu(\theta)} \mu(dx)$$

is called the NEF generated by μ.

The two special cases of the geometric and exponential families have the following NEF features:

- Geometric:

$$\mu(dx) = \sum_{k=0}^{\infty} \delta_k(dx), \; L_\mu(\theta) = (1 - e^\theta)^{-1}, k_\mu(\theta) = -\ln(1 - e^\theta), \Theta(\mu) = (-\infty, 0),$$

where δ_k is the Dirac mass on k. In this case

$$P(\theta, \mu)(dx) = \sum_{x \in \mathbb{N}_0} (1 - q) q^x \delta_x$$

where $q = e^{\theta} < 1$. Let $X_1, ..., X_n$ be i.i.d. r.v.'s with common geometric distribution with parameter q, then the p.d.f. of $Y = \min(X_1, ..., X_n)$ is geometric with parameter q^n, or in its NEF p.d.f. form with $\theta \longmapsto n\theta$.

- Exponential:

$$\mu(dx) = \mathbf{1}_{(0,\infty)}(dx), L_\mu(\theta) = (-\theta)^{-1}, k_\mu(\theta) = -\ln(-\theta), \Theta(\mu) = (-\infty, 0),$$

where its known p.d.f. form is

$$\lambda e^{-\lambda x}\mathbf{1}_{(0,\infty)}, \lambda > 0,$$

in which case

$$\theta = -\lambda.$$

If $X_1, ..., X_n$ be i.i.d. r.v.'s with common exponential distribution with parameter λ then the p.d.f. of $Y = \min(X_1, ..., X_n)$ is again exponential with parameter $n\lambda$, or in its NEF p.d.f. form with $\theta \longmapsto n\theta$.

3 The main result

Theorem 3.1. *Let $\mu \in \mathcal{M}(\mathbb{R})$ and $n \geq 2$ be an an integer. Let $X_1, ..., X_n$ be i.i.d. r.v's with common distribution $P(\theta, \mu)$ and denote by Q_θ the distribution of $Y = \min(X_1, ..., X_n)$. Then there exist a measure $\nu \in \mathcal{M}(\mathbb{R})$, an NEF $F(\nu)$ and a differentiable mapping $\theta \mapsto \alpha(\theta)$ from $\Theta(\mu)$ to $\Theta(\nu)$ such that $Q_\theta = P(\alpha(\theta), \nu)$ for all $\theta \in \Theta(\mu)$ if and only if $F(\mu)$ is a positive affine transformation of either the NEF of geometric distributions or the NEF of exponential distributions.*

Proof. The statement $\Leftarrow$ is simple as can be seen from the remarks at the end of Section 2. Indeed, with the choices of μ made there, we have for both, the geometric and exponential cases, that $\mu = \nu$ and $\alpha(\theta) = n\theta$.

We prove the statement $\Rightarrow$ in six steps. In the first step we derive the functional equation (3.3) which provides a necessary condition for $Q_\theta \sim Y = \min(X_1, ..., X_n)$ to belong to some NEF $F(\nu)$. The second step proves that the support of μ is bounded on the left, while the third step shows that such a support is unbounded on the right. The fourth step further analyzes the functional equation (3.3) and provides an important equation (3.7) associated with the measure μ. More specifically the problem is then being reduced to the case where the support interval (i.e., the convex hull of the support) of μ is exactly $[0, \infty)$. If we denote by μ_x the translate of $\mu(dt)$ by $t \mapsto t - x$ and then truncate at zero, the equality (3.7) is $k'_{\mu_x} = k'_\mu$ for μ almost all x. This equality reduces the characterization problem to the problem of whether μ possesses at least one atom or not. If μ has at least one atom the the fifth step proves that μ generates the geometric NEF. Otherwise, the sixth step shows that μ generates the exponential NEF. Such six steps then conclude the proof.

First step. This step is devoted to the setting of the functional equation (3.3) below. For simplicity, we write $k = k_\mu, \Theta = \Theta(\mu)$ and so on. In the sequel we write

$$\int_{a^-}^{b^+} f(t)\mu(dt) \text{ for } \int_{[a,b]} f(t)\mu(dt) \text{ and } \int_{a^+}^{b^+} f(t)\mu(dt) \text{ for } \int_{(a,b]} f(t)\mu(dt).$$

If the law of Y belongs to an NEF $F(\nu)$ then for $\theta \in \Theta$ and real y, the number $P(Y \geq y)$ can be represented in two different ways, by which one gets the following equality

$$e^{-nk(\theta)} \left(\int_{y^-}^{\infty} e^{\theta t}\mu(dt) \right)^n = e^{-k_\nu(\alpha(\theta))} \int_{y^-}^{\infty} e^{\alpha(\theta)t}\nu(dt),$$

and hence the following equality, between two probability measures, holds:

$$ne^{-nk(\theta)}\left(\int_{y^-}^{\infty} e^{\theta t}\mu(dt)\right)^{n-1} e^{\theta y}\mu(dy) = e^{-k_\nu(\alpha(\theta))}e^{\alpha(\theta)y}\nu(dy).$$

This proves that the measures ν and μ are equivalent and we can introduce the Radon Nikodym derivative $g(y) = \frac{d\nu}{d\mu}(y)$. Hence, the following equality which holds μ almost everywhere:

$$ne^{-nk(\theta)+k_\nu(\alpha(\theta))}\left(\int_{y^-}^{\infty} e^{\theta t}\mu(dt)\right)^{n-1} e^{(\theta-\alpha(\theta))y} = g(y).$$

By denoting $g_n(y) = \left(\frac{g(y)}{n}\right)^{1/(n-1)}$ and

$$A(\theta) = \frac{-\theta+\alpha(\theta)}{n-1}, \quad B(\theta) = \frac{nk(\theta) - k_\nu(\alpha(\theta))}{n-1}, \tag{3.1}$$

and elevating to the power $1/(n-1)$, we get the following equality which holds μ almost everywhere:

$$e^{-yA(\theta)-B(\theta)}\int_{y^-}^{\infty} e^{\theta t}\mu(dt) = g_n(y). \tag{3.2}$$

Assume, without loss of generality, that μ and ν are probability measures. Then, the Hölder inequality, applied to the pair of functions $(g,1)$ and to $(p,q) = (n-1,(n-1)/(n-2))$, shows that $\int_{-\infty}^{\infty} g_n(y)\mu(dy) < \infty$. Integrating (3.2) on $[x,\infty)$ with respect to $\mu(dy)$ yields for all $\theta\in\Theta$

$$\int_{x^-}^{\infty}\left(e^{-yA(\theta)-B(\theta)}\int_{y^-}^{\infty} e^{\theta t}\mu(dt)\right)\mu(dy) = \int_{x^-}^{\infty} g_n(y)\mu(dy).$$

Now, by differentiating, with respect to θ, of both sides of the latter equality, we obtain

$$\int_{x^-}^{\infty}\left(e^{-yA(\theta)-B(\theta)}\int_{y^-}^{\infty} e^{\theta t}(t - yA'(\theta) - B'(\theta))\mu(dt)\right)\mu(dy) = 0.$$

Since the latter equality holds for all x, it follows that for each fixed $\theta\in\Theta$,

$$\int_{x^-}^{\infty} e^{\theta t}(t - xA'(\theta) - B'(\theta))\mu(dt) = 0, \tag{3.3}$$

which holds $\mu(dx)$ almost everywhere. The equality (3.3) holds in particular for any element x of the support S of the measure μ. To prove this statement, we denote by $H(x)$ the left hand side of (3.3). Then locally, H has a bounded variation (i.e., it is the difference of two non-increasing functions) and its discontinuity points are the atoms of μ. Therefore $H(x) = 0$ if x is an atom of μ. If $x\in S$ and is not an atom of μ then there exists a sequence (x_k) such that $H(x_k) = 0$ for all k and such that $x_k \to x$. Since H is continuous in x it follows that $H(x) = 0$ for all $x\in S$.

Second step. We prove that the support of μ is bounded on the left. If not, the equality (3.3) holds for some fixed $\theta\in\Theta$ and for some sequence (x_k) such that $\lim_{k\to\infty} x_k = -\infty$. This implies that $A'(\theta) = 0$ and $B'(\theta) = k'(\theta)$. But then clearly the equality

$$\int_{x_k^-}^{\infty} e^{\theta t}(t - k'(\theta))\mu(dt) = 0$$

cannot hold for all k. Indeed, if k_0 is such that $x_{k_0} \le k'(\theta)$ then such an equality would imply that for any $k > k_0$

$$0 = \int_{x_k^-}^{x_{k_0}^-} e^{\theta t}(t - k'(\theta))\mu(dt),$$

while the right hand side is negative for k large enough.

Third step. This step proves that the support of μ is unbounded on the right. It relies on the following lemma, which has its own interest with its characterisation of the distribution $B(1,a)$ up to a dilation by b :

Lemma 1. Let P be a non-Dirac probability on $[0,\infty)$ and $K>0$ such that for P almost all x we have

$$\int_{0_-}^{x^+} tP(dt) = Kx\int_{0_-}^{x^+} P(dt). \tag{3.4}$$

Then $K<1$ and there exists $b>0$ such that $P(dt)=\frac{a}{b^a}t^{a-1}1_{(0,b)}(t)dt$, where $a=K/(1-K)$.

Proof. If $K>1$ then for at least one $x>0$ we have

$$\int_{0^-}^{x^+} P(dt) = \frac{1}{K}\int_{0^-}^{x^+}\frac{t}{x}P(dt) < \int_{0^-}^{x^+} P(dt),$$

which is a contradiction. If $K=1$ then $0=\int_{0^-}^{x^+}(t-x)P(dt)$ for P almost all x. This implies that $t-x=0$ for $P(dt)P(dx)$ almost all (t,x), which is possible only if P is a Dirac measure, a contradiction. The probability measure P has no atom on $t_0>0$ since (3.4) implies $t_0P(\{t_0\})=Kt_0P(\{t_0\})$ which contradicts that $K<1$. Similarly, P has no atom on zero. If not, since for at least one $x>0$ one has

$$\int_{0_+}^{x} P(dt) \geq \int_{0_+}^{x}\frac{t}{x}P(dt) = \int_{0_-}^{x}\frac{t}{x}P(dt) = KP(\{0\}) + \int_{0_+}^{x} P(dt) > \int_{0_+}^{x} P(dt),$$

we get a contradiction.

The support S of P contains 0. If not, there exists b in S such that $P([0,b))=0$. Since P is not Dirac there exists a sequence $x_n\searrow b$ such that

$$\int_{b}^{x_n}\frac{t}{x_n}P(dt) = A\int_{b}^{x_n} P(dt).$$

Now consider the conditional probability P_n which is $P(dt)$ conditioned on $b<t<x_n$. Then, P_n converges weakly to δ_b (the simplest way to prove this is to use the distribution function of P_n). Since $\int_b^{x_n} tP_n(dt)=Kx_n$, then by passing to the limit we get the contradiction for $K=1$.

The support S of P is an interval containing zero. If not, and since $0\in S$, there exist $0<x_1<x_2$ such that $P((x_1,x_2))=0$, $x_1,x_2\in S$ and $\int_0^{x_1}P(dt)>0$. Hence, from (3.4), we get the following contradiction

$$Kx_1\int_0^{x_1}P(dt) \overset{(a)}{=} \int_0^{x_1}tP(dt) \overset{(b)}{=} \int_0^{x_2}tP(dt) \overset{(c)}{=} Kx_2\int_0^{x_2}P(dt) \overset{(d)}{=} Kx_2\int_0^{x_1}P(dt),$$

where the equalities (a) and (c) stem from (3.4) and the fact that x_1 and x_2 are in S. The equalities (b) and (d) come from the fact that $P((x_1,x_2))=0$.

Now, since P has no atoms, the function

$$f(x)=\int_0^x tP(dt) - Kx\int_0^x P(dt)$$

is continuous. Furthermore, f is zero P almost everywhere. This implies that f is zero on the support S of P. If not, there exists $x_0\in S$ such that $|f(x_0)|>0$ and an open interval (x_0-h,x_0+h) such that $|f(x)|>0$ if $|x-x_0|<h$. However,

$$\int_{x_0-h}^{x_0+h}|f(t)|\mu(dt)=0$$

and thus $(x_0 - h, x_0 + h)$ and S are disjoint (recall that S is the complementary set of the largest open set with P zero measure). Hence, a contradiction follows.

Accordingly, $S = [0, b]$ for some real b or $S = [0, \infty)$. Denote by S_0 the interior of S. We have seen that for all $x \in S$, $f(x) = 0$. We rewrite this fact as

$$\int_0^x tP(dt) = Kx \int_0^x P(dt).$$

Differentiating this equality (in the Stieltjes sense) we get (on S_0) that

$$xP(dx) = a\left(\int_0^x P(dt)\right) dx,$$

where $a = \frac{K}{1-K}$. This shows that $P(dx) = g(x)dx$ is absolutely continuous. In fact, from $xg(x) = a \int_0^x g(t)dt$, it follows that the function g is continuous and even differentiable on S_0. This leads to the differential equation $g'(x)/g(x) = (a-1)/x$ on S_0 and $g(x) = Cx^{a-1}$, where $C > 0$. If S is unbounded then g cannot be a probability density. Therefore $S = [0, b]$ is bounded and the lemma is proved. ■

We now prove the claim of Step 3 that the support of μ is unbounded on the right. If not, from Step 2, we may assume without loss of generality that the support interval of μ is exactly $[0, b]$ with $b > 0$. Substituting $x = 0$ in (3.3) gives $B'(\theta) = k'(\theta)$. We now show that $A'(\theta) = 1 - \frac{1}{b}k'(\theta)$. To see this we rewrite (3.3) as follows

$$\frac{\int_{x_-}^{b^+} e^{\theta t} t\mu(dt)}{\int_{x_-}^{b^+} e^{\theta t}\mu(dt)} = xA'(\theta) + k'(\theta) \tag{3.5}$$

and we do $x \nearrow b$ in (3.5). The left hand side converges to b and $A'(\theta) = 1 - \frac{1}{b}k'(\theta)$ is proved. This now leads to the equation

$$\frac{\int_{x^-}^{b^+} e^{\theta t}(t - x)\mu(dt)}{\int_{x^-}^{b^+} e^{\theta t}\mu(dt)} = \left(1 - \frac{x}{b}\right) k'(\theta). \tag{3.6}$$

Fix θ, consider the change of variable $t \mapsto b - t$ and apply Lemma 1 to the image $P(dt)$ of the probability $e^{\theta t - k(\theta)}\mu(dt)$ and to $A = k'(\theta)/b$. Then, it follows that the a of Lemma 1 is $a(\theta) = k'(\theta)/(b - k'(\theta))$. Since the support interval of P is also $[0, b]$ we can claim that

$$e^{\theta t - k(\theta)}\mu(dt) = a(\theta)(b - t)^{a(\theta)-1} 1_{(0,b)}(t)dt,$$

an equality which cannot hold for all θ. One may realize this as follows. Since

$$\mu(dt) = e^{-t\theta + (a(\theta)-1)\log(b-t) + c(\theta)} 1_{(0,b)}(t)dt,$$

where $c(\theta) = k(\theta) + \log a(\theta)$, we have, by differentiating by θ, that for all $(\theta, t) \in \Theta \times (0, b)$,

$$-t + a'(\theta)\log(b - t) + c'(\theta) = 0.$$

Then, differentiating by t, we get $b - t = a'(\theta)$, which is clearly impossible.

Fourth step. From Steps 2 and 3, we may assume throughout the sequel that the support interval of μ is exactly $[0, \infty)$. This assumption implies that we are allowed to substitute $x = 0$ in (3.3) to obtain

$$\int_{0^-}^{\infty} e^{\theta t}(t - B'(\theta))\mu(dt) = 0,$$

which shows that $B' = k'$. By the definition of B in (3.1), this implies that $k(\theta) - k_\nu(\alpha(\theta))$ is a constant. We denote by $\mu_x(du)$ the image of the measure μ by the map $t \longmapsto u = t - x$ multiplied by the function $\mathbf{1}_{[0,\infty)}(u)$. The equality (3.3) can then be reformulated as

$$k'_{\mu_x}(\theta) = k'(\theta), \tag{3.7}$$

for $\mu(dx)$ almost everywhere. We now analyze (3.7) according to whether μ has at least one atom (Fifth Step), an assumption that will lead to the geometric NEF, or not (Sixth Step), a fact that will lead to the exponential NEF.

Fifth step. Assume that μ has an atom x_0. We prove that there exists a countable additive subgroup G of $\mathbb{R}$ and a real character χ of G such that

$$\mu(dt) = \mu(0) \sum_{x \in G \cap [0,\infty)} e^{\chi(x)} \delta_x(dt),$$

where $\mu(x)$ denotes the mass of the atom x.

This assumption implies that (3.7) is true for $x = x_0$ and thus that μ has an atom on 0 (and thus are all the measures μ_x for which (3.7) is true). This implies that μ is purely atomic. Denote by S the set of atoms of μ. From (3.7) we infer that for all $x \in S$ we have

$$S = (S - x) \cap [0, \infty).$$

Denote $G = S \cup (-S)$. Then G is an additive group with $S = G \cap [0, \infty)$. Write $\mu(dt) = \sum_{x \in S} \mu(x)\delta_x(dt)$, then (3.7) implies that for all $x \in S$ we have

$$\mu_x(dt) = \frac{\mu(x)}{\mu(0)} \mu(dt).$$

Calculating the mass of this measure on $s \in S$ we get

$$\mu(s) = \frac{\mu(0)}{\mu(x)} \mu_x(s) = \frac{\mu(0)}{\mu(x)} \mu(x + s).$$

For $x \in S$, denote $\chi(x) = \log \mu(x) - \log \mu(0)$ and for $x \in -S$ denote $\chi(x) = -\chi(-x)$. Then the latter equality implies

$$\chi(x + s) = \chi(x) + \chi(s),$$

that is, χ is a real character of G.

We now prove that G is $a\mathbb{Z}$ for some for some $a > 0$. If not, then G is a dense in $\mathbb{R}$. Then, either any pair (x, x') of $G \backslash \{0\}$ is such that x/x' is rational, or there exists a pair such that x/x' is irrational. Without loss of generality, we may assume for the latter two cases that $1 \in G$. In the first case (where x/x' is rational) there exist arbitrary small rational numbers $x \in G$ such that $\chi(x) = x\chi(1)$. Thus, for $A > 0$, the family $\left\{e^{\chi(x)} : x \in G \cap [0, A]\right\}$ cannot be summable and μ is not a Radon measure. Similarly, for the second case (x/x' is irrational), G contains a subgroup $\mathbb{Z}(\alpha)$ for some irrational number α (where $\mathbb{Z}(\alpha)$ is the set of $a + b\alpha$ with a, b in $\mathbb{Z}$). By denoting $p_1 = e^{\chi(1)}$ and $p_2 = e^{\chi(\alpha)}$ we obtain that $p_1^a p_2^b = e^{\chi(a+b\alpha)}$. We now need to prove that

$$\sum \left\{p_1^a p_2^b : 0 \le a + b\alpha \le A\right\} = \infty. \tag{3.8}$$

This can be accomplished by a tedious discussion and analysis of the nine cases $0 < p_1 < 1, p_1 = 1$ and $p_1 > 1$ combined with $0 < p_2 < 1, p_2 = 1$ and $p_2 > 1$ (we omit details for brevity). This, however, would finally show that μ cannot be a Radon measure.

Thus we conclude the case where μ has at least one atom by stating that for this case there exist $a > 0$ and numbers $p = e^{\chi(a)} > 0$ and $q = \mu(0)$ such that

$$\mu(dt) = \mu(0) \sum_{n=0}^{\infty} qp^n \delta_{na}(dt).$$

This is equivalent to saying that $F(\mu)$ is the image of the geometric distributions by the dilation $n \longmapsto an$.

Sixth step. We assume that μ has no atoms. Denote by $X \subset [0,\infty)$ the set of x such that (3.7) holds. We prove that the closure $\bar{X}$ of X is the support S of μ. To see that $S \subset \bar{X}$, we choose $x_0 \in S$. If there is no sequence (x_n) of X converging to x_0, this would imply the existence of $\epsilon > 0$ such that $\mu([x_0-\epsilon, x_0+\epsilon]) = 0$ and thus contradict the fact that $x_0 \in S$. To see that $X \subset S$ we choose $x_0 \in X$. If $x_0 \notin S$ then this would imply the existence of $\epsilon > 0$ such that $\mu([x_0-\epsilon, x_0+\epsilon]) = 0$. Since $0 \in S$, the measure μ_{x_0} cannot be equivalent to μ. Thus, the statement that $S = \bar{X}$ is proved.

Now, the fact that μ has no atoms implies that $x \longmapsto \mu_x$ is a continuous function on $\mathbb{R}$ for the vague topology of Radon measures. The equality (3.7) is thus equivalent to the existence of a function χ on X such that

$$\mu_x(dt) = e^{\chi(x)}\mu(dt), \tag{3.9}$$

and the preceding remark implies that χ is a continuous function on X and is extendable in a continuous function to $\bar{X}$. Thus (3.7) and (3.8) hold on S. Now we observe that (3.7) implies that for all $x \in S$ we have $S = (S-x) \cap [0,\infty)$. Thus $G = S \cup (-S)$ is an additive subgroup of $\mathbb{R}$. Since G is closed, then either $G = \{0\}$, or there exists $a > 0$ such that $G = a\mathbb{Z}$ or $G = \mathbb{R}$. Such two cases can be excluded since μ has no atoms, and thus we get $S = [0,\infty)$.

We now show that $\chi(x+s) = \chi(x) + \chi(s)$ for all $x \geq 0$ and $s \geq 0$. For this we observe that (3.7) implies that for all $x \geq 0$ the measure μ_x generates the NEF $F(\mu)$. Thus μ_x must share with μ the property (3.7), and for $s \geq 0$ we therefore have

$$\mu_{x+s}(dt) = e^{\chi(x)}\mu(dt).$$

Since μ_x and μ are proportional, the factor $e^{\chi(x)}$ is the same. Since we also have $\mu_{x+s}(dt) = e^{\chi(x+s)}\mu_x(dt)$, the equality $\chi(x+s) = \chi(x) + \chi(s)$ follows.

As χ is continuous, it is simple to see that there exists $b \in \mathbb{R}$ such that $\chi(x) = bx$. One can consult Bingham, Teugels and Goldie for reference to this Cauchy functional equation. By introducing the measure $\tilde{\mu}(dt) = e^{-bt}\mu(dt)$, we have $F(\tilde{\mu}) = F(\mu)$. Furthermore (3.9) implies that for all $x \geq 0$ we have

$$\tilde{\mu}_x(dt) = \tilde{\mu}(dt).$$

This implies that for all intervals $I \subset [0,\infty)$, we have $\tilde{\mu}(x+I) = \tilde{\mu}(I)$. Thus $\tilde{\mu}$ is proportional to the restriction of the Lebesgue measure to $[0,\infty)$ and the theorem is proved. □

Acknowledgement. We thank the referee for a very careful reading of the original version of the paper.

References

[1] Balkema, A.A. and de Haan, L. (1974). Residual life time at great age. *Ann. Probab.* **2**(5), 792–804. MR-0359049

[2] Bar-Lev, S.K. and Bshouty, D. (2008). Exponential families are not preserved by the formation of order statistics. *Statist. Probab. Lett.* **78**, 2787–2792. MR-2465123

[3] Bingham, N.H., Goldie, C.M. and Teugels, J.L (1989) *Regular Variation.* Cambridge University Press. MR-1015093

[4] Letac, G. and Mora, M. (1990). Natural real exponential families with cubic variance functions. *Ann. Statist.* **18**, 1–37. MR-1041384

[5] Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131. MR-0423667