PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

CHARACTERIZATIONS OF EXPONENTIAL DISTRIBUTION VIA CONDITIONAL EXPECTATIONS OF RECORD VALUES

George P. Yanev

We prove that the exponential distribution is the only one which satisfies a regression identity. This identity involves conditional expectation of the sample mean of record values given two record values outside of the sample.

1. Discussion of results

Before formulating and discussing the obtained results, let us introduce some notation regarding record values. Let $X_1, X_2, \ldots$ be independent copies of a random variable X with absolutely continuous cumulative distribution function (c.d.f.) $F(x)$. (Upper) Record value is an observation in a discrete time series, which exceeds all previous observations, i.e., X_j is a (upper) record value if $X_j > X_i$ for all $i < j$. Define the sequence $\{T_n, n \geq 1\}$ of record times by $T_1 = 1$ and $T_n = \min\{j : X_j > X_{T_{n-1}}, j > T_{n-1}\}$ for $n > 1$. Then the corresponding record values are $R_n = X_{T_n}$ for $n = 1, 2, \ldots$ (see Nevzorov (2001)).

Let $F(x)$ be the exponential c.d.f. given by

$$(1) \qquad\qquad F(x) = 1 - e^{-cx}, \qquad (x \geq 0),$$

where $c > 0$. It is well-known (cf. Arnold et al. (1998), Section 4.2.2), that a single regression function (conditional expectation) of the form $E(R_{n+1}|R_n)$ is

enough to characterize a continuous c.d.f. Several authors have studied characterizations of exponential distributions in terms of the regression of one record value with two other record values as covariates, i.e., regression functions given for $1 \leq k \leq n-1$ and $r \geq 1$ by

(2) $E[g(R_n)|R_{n-k} = u, R_{n+r} = v] = \varphi(u, v), \qquad (0 \leq u < v),$

where the function g satisfies certain regularity conditions. Let $\bar{g}(u, v)$ denote the average value of the function $g(x)$ over the interval $[u, v]$, i.e.,

$$\bar{g}(u, v) = \frac{1}{v - u} \int_u^v g(t)dt.$$

Bairamov et al. (2005) study the particular case of (2) when both covariates are adjacent (only one spacing away) to R_n. They prove, under some regularity conditions on F and g and using different notation, that F is exponential if and only if

$$E\left[g(R_n)\middle|R_{n-1} = u, R_{n+1} = v\right] = \bar{g}(u, v), \qquad (0 \leq u < v).$$

On the other hand, Yanev et al. (2008) prove the following necessary condition for a continuous c.d.f. to be exponential. If F is exponential then for $1 \leq k \leq n-1$ and $r \geq 1$

(3) $E\left[\dfrac{g^{(k-1+r-1)}(R_n)}{k+r-1}\middle|R_{n-k} = u, R_{n+r} = v\right]$
$$= \binom{k-1+r-1}{k-1} \frac{\partial^{k-1+r-1}}{\partial u^{r-1}\partial v^{k-1}} \bar{g}(u, v),$$

where $0 \leq u < v$. Let us demonstrate (3) using two particular choices of the function $g(x)$. If $g(x) = x^{k+r-1}/(k+r-1)!$, then (3) becomes

(4) $E[R_n|R_{n-k} = u, R_{n+r} = v] = \dfrac{ru + kv}{k+r}, \qquad (0 \leq u < v).$

Note that the right-hand side of (4) is the weighted mean of the covariates' values. The weight of one covariate is proportional to the number of spacings R_n is away from the other covariate. As a result, the closer R_n is to one of the covariates, the more affected its value is from this covariate. Another choice of the function $g(x)$ could be $g(x) = (-1)^{k+r-2}x^{-2}/(k+r-1)!$ when (3) is equivalent to

$$E\left[\frac{1}{R_n^{k+r}}\middle|R_{n-k} = u, R_{n+r} = v\right] = \frac{1}{u^r v^k}, \quad (0 \leq u < v).$$

Whether (3) is not only a necessary, but also a sufficient condition for an absolutely continuous F to be exponential for any $1 \leq k \leq n-1$ and $r \geq 1$ is an open question. Yanev et al. (2008) obtain an affirmative answer when $r = 1$, i.e., one of the two covariates is adjacent to R_n. They show, under some regularity assumptions on F and g, that $F(x)$ is exponential if and only if

$$(5) \quad E\left[\frac{g^{(k-1)}(R_n)}{k}\bigg|R_{n-k}=u, R_{n+1}=v\right] = \frac{\partial^{k-1}}{\partial v^{k-1}}\,\bar{g}(u,v), \qquad (0 \leq u < v).$$

A different direction for extending the characterization results above is to consider conditional expectations of sums of record values. Akhundov and Nevzorov (2007) prove that an absolutely continuous F is exponential if and only if

$$(6) \qquad E\left[\frac{R_2 + R_3 + \ldots + R_n}{n-1}\bigg|R_1 = u, R_{n+1} = v\right] = \frac{u+v}{2}.$$

Our first result is one generalization of (6).

Theorem 1. *Let $n \geq 2$ be an integer. Suppose $F(x)$ is absolutely continuous, $g(x)$ is continuous in $[0,\infty)$, and $g(v) \neq \bar{g}(0,v)$ for every $v > 0$. Then (1) holds if and only if for $0 \leq u < v < \infty$*

$$(7) \qquad E\left[\frac{g(R_2) + g(R_3) + \ldots + g(R_n)}{n-1}\bigg|R_1 = u, R_{n+1} = v\right] = \bar{g}(u,v).$$

Next, we extend Theorem 1 to the case when the second covariate is two spacings away from R_n. The general case of non-adjacent covariates remains an open problem. Denote, for a continuous function $g(x)$ and $n = 1, 2, \ldots$

$$(8) \qquad I_n(u,v) = \int_0^1 g(u + (v-u)z)z^{n-1}dz, \qquad (0 \leq u < v).$$

Note that $I_1(u,v) = \bar{g}(u,v)$. Denote $H(x) = -\ln(1 - F(x))$ for $x \geq 0$ and $h(x) = H'(x)$, i.e., $H(x)$ is the cumulative hazard function of X and $h(x)$ is its hazard (failure) rate.

Theorem 2. *Suppose $F(x)$ satisfies the following conditions.*
(i) $F(x)$ is absolutely continuous in $[0,\infty)$ and $F''(x)$ is continuous in $(0,\infty)$;
(ii) $h(x)$ is nowhere constant in a small interval $(0,\varepsilon)$ for $\varepsilon > 0$.

Assume that $g(x)$ is continuous in $[0, \infty)$, $g'(x)$ is continuous in $(0, \infty)$, and $g'(0+) \neq 0$. Let $n \geq 2$ be an integer. Then (1) holds if and only if for $0 \leq u < v < \infty$

$$(9) \quad E\left[\frac{g(R_2) + g(R_3) + \ldots + g(R_n)}{n-1} \Big| R_1 = u, R_{n+2} = v\right]$$

$$= \bar{g}(u, v) - \frac{nI_n(u, v) - \bar{g}(u, v)}{n-1}.$$

Setting in Theorem 2, $g(x) = x$ and thus $nI_n(u, v) = (u + nv)/(n+1)$, and taking into account (9) we obtain the following extension of (6).

Corollary 1. *Suppose $F(x)$ satisfies the following conditions.*
(i) $F(x)$ is absolutely continuous in $[0, \infty)$ and $F''(x)$ is continuous in $(0, \infty)$;
(ii) $h(x)$ is nowhere constant in a small interval $(0, \varepsilon)$ for $\varepsilon > 0$. Let $n \geq 2$ be an integer. Then (1) holds if and only if for $0 \leq u < v < \infty$

$$(10) \quad E\left[\frac{R_2 + R_3 + \ldots + R_n}{n-1} \Big| R_1 = u, R_{n+2} = v\right] = \frac{u+v}{2} - \frac{v-u}{2(n+1)}$$

$$= \frac{(n+2)u + nv}{n+2+n}.$$

Notice that (4) implies that a necessary condition for F to be exponential is

$$(11) \quad E\left[\frac{R_2 + R_3 + \ldots + R_n}{n-1} \Big| R_1 = u, R_{n+m} = v\right]$$

$$= \frac{(n+2m-2)u + nv}{n+2m-2+n}, \qquad m \geq 1.$$

The results (6) and Corollary 2 prove the sufficiency of (11) for $m = 1$ and $m = 2$.

2. Proofs

Lemma 1. *Let r and $n \geq 2$ be positive integers. If (1) holds, then*

$$(12) \quad E\left[\sum_{j=2}^{n} g(R_j) \mid R_1 = u, R_{n+r} = v\right]$$

$$= (N+1) \int_0^1 g(u + (v-u)z) \left[1 - \sum_{j=n-1}^{N} \binom{N}{j} z^j (1-z)^{N-j}\right] dz,$$

where $N = n + r - 3$. As usual we assume $\sum_j^i (\cdot) = 0$ when $i < j$.

Proof. It is well-known (e.g., Nevzorov (2001), p.11 and p.69) that if (1) holds, then $R_n = \sum_{j=1}^{n} \xi_j$, where $\{\xi_j\}$ are i.i.d. exponential with (1). Therefore, for $2 \leq j \leq n$

$$
\begin{aligned}
E\left[R_j | R_1 = u, R_{n+r} = v\right] \\
= \quad & E\left[\xi_1 + \xi_2 + \ldots + \xi_j | \xi_1 = u, \xi_1 + \xi_2 + \ldots + \xi_{n+r} = v\right] \\
= \quad & E\left[u + \xi_2 + \ldots + \xi_j | \xi_2 + \ldots + \xi_{n+r} = v - u\right] \\
= \quad & E\left[u + Y_j | Z_{n+r} = v - u\right], \qquad \text{say.}
\end{aligned}
$$

We obtain the following density functions related to Y_j and Z_{n+r} for $2 \leq j \leq n$

$$
f_{Y_j, Z_{n+r}}(y, z) = \frac{c^{n+r-1} y^{j-2} (z-y)^{n+r-j-1}}{(j-2)!(n+r-j-1)!} e^{-cz}, \quad f_{Z_{n+r}}(z) = \frac{c^{n+r-1}}{(N+1)!} z^{N+1} e^{-cz},
$$

$$
f_{Y_j | Z_{n+r}}(y|z) = \frac{(N+1)!}{(j-2)!(n+r-j-1)!} \left(\frac{y}{z}\right)^{j-2} \left(1 - \frac{y}{z}\right)^{n+r-j-1} \frac{1}{z},
$$

where $N = n + r - 3$. Now we are in a position to prove (7). Indeed,

$$
\begin{aligned}
E\left[\sum_{j=2}^{n} g(R_j) | R_1 = u, R_{n+r} = v\right] &= E\left[\sum_{j=2}^{n} g(u + Y_j) | Z_{n+r} = v - u\right] \\
= \quad & \int_{0}^{v-u} g(u+t) \sum_{j=2}^{n} \frac{(N+1)!}{(j-2)!(n+r-j-1)!} \left(\frac{t}{v-u}\right)^{j-2} \left(1 - \frac{t}{v-u}\right)^{n+r-j-1} \frac{dt}{v-u} \\
= \quad & (N+1) \int_{0}^{1} g(u + (v-u)z) \sum_{j=2}^{n} \binom{N}{j-2} z^{j-2}(1-z)^{n+r-j-1} dz \\
= \quad & (N+1) \int_{0}^{1} g(u + (v-u)z) \sum_{i=0}^{n-2} \binom{N}{i+k-1} z^{i+k-1}(1-z)^{N-(i+k-1)} dz,
\end{aligned}
$$

which is equivalent to (12). $\qquad \square$

2.1. Proof of Theorem 1

Recall the formula for the conditional density $f_{j|1,n+r}(t|u,v)$, say, of R_j given $R_1 = u$ and $R_{n+r} = v$, where $r \geq 1$ and $2 \leq j \leq n$ (e.g., Ahsanullah (2004), p.6).

It is given for $u < t < v$ and $2 \le j \le n$ by

(13) $\quad f_{j|1,n+r}(t|u,v)$

$$= \frac{(n+r-2)!}{(j-2)!(n+r-1-j)!} \left[\frac{H(t)-H(u)}{H(v)-H(u)} \right]^{j-2} \left[\frac{H(v)-H(t)}{H(v)-H(u)} \right]^{n+r-1-j} \frac{h(t)}{H(v)-H(u)}.$$

   Sufficiency. Assuming (7), we will prove that $F(x)$ is exponential (1). Using (13) with $r = 1$ and the binomial formula, we obtain

(14) $$\sum_{j=2}^{n} f_{j|1,n+1}(t|u,v) = \frac{(n-1)h(t)}{H(v)-H(u)}.$$

Taking into account (14), for the left-hand side of (7) we have

$$E\left[\sum_{j=2}^{n} g(R_j)|R_1 = u, R_{n+1} = v \right] \quad = \quad \int_u^v \sum_{j=2}^{n} g(t) f_{j|1,n+1}(t|u,v) \; dt$$

$$= \quad \frac{n-1}{H(v)-H(u)} \int_u^v g(t) dH(t)$$

and thus (7) is equivalent to

$$\frac{1}{H(v)-H(u)} \int_u^v g(t) dH(t) = \frac{1}{v-u} \int_u^v g(t) dt.$$

Letting $u \to 0+$, and hence $H(u) \to 0$, we obtain

$$\int_0^v g(t) dH(t) = \frac{H(v)}{v} \int_0^v g(t) dt.$$

Differentiating both sides with respect to v, after rearranging terms, we find

$$[g(v) - \bar{g}(0,v)]h(v) = \frac{H(v)}{v}[g(v) - \bar{g}(0,v)].$$

Dividing both sides by $g(v) - \bar{g}(0,v) \ne 0$, we obtain the equation

$$\frac{h(v)}{H(v)} = \frac{1}{v},$$

which only solution is the exponential $F(x)$ given by (1). Indeed, integrating both sides with respect to v, we obtain

$$\ln H(x) = \ln v + \ln c \qquad (c > 0)$$

and thus $H(x) = -\ln(1 - F(x)) = cx$ and (1) follows.

Necessity. If $F(x)$ is exponential, then (7) follows from the lemma setting $r = 1$.

2.2. Proof of Theorem 2

Sufficiency. Recall the notation $f_{j|1,n+r}(t|u,v)$ introduced before (13). Now, (13) with $r = 2$ and the binomial formula imply

$$\sum_{j=2}^{n} f_{j|1,n+2}(t|u,v)$$

$$= \frac{nh(t)}{(H(v) - H(u))^n} \sum_{j=2}^{n} \binom{n-1}{j-2} (H(t) - H(u))^{j-2}(H(v) - H(t))^{n-j+1}$$

$$= \frac{nh(t)}{(H(v) - H(u))^n} \sum_{i=0}^{n-2} \binom{n-1}{i} (H(t) - H(u))^{i}(H(v) - H(t))^{n-1-i}$$

$$(15) \qquad = \frac{nh(t)}{H(v) - H(u)} \left[1 - \left(\frac{H(t) - H(u)}{H(v) - H(u)} \right)^{n-1} \right].$$

It follows from (15) that

$$(16) \qquad E\left[\sum_{j=2}^{n} g(R_j) | R_1 = u, R_{n+2} = v \right] = \int_{u}^{v} g(t) \sum_{j=2}^{n} f_{j|1,n+2}(t|u,v)dt$$

$$= \frac{n}{H(v) - H(u)} \int_{u}^{v} g(t) \left[1 - \left(\frac{H(t) - H(u)}{H(v) - H(u)} \right)^{n-1} \right] dH(t).$$

It follows from (9) and (16), letting $u \to 0$ (and hence $H(u) \to 0$) and multiplying by $H^n(v)(n-1)/n$ that

$$\int_{0}^{v} g(t) \left[H^{n-1}(v) - H^{n-1}(t) \right] dH(t) = H^n(v)[\bar{g}(v) - I_n(v)],$$

where $\bar{g}(v) = \bar{g}(0, v)$ and $I_n(v) = I_n(0, v)$. Differentiating with respect to v and dividing by $H^{n-2}(v)h(v)$ we obtain

$$(n-1)\int_0^v g(t)dH(t) = nH(v)[\bar{g}(v) - I_n(v)] + \frac{H^2(v)}{h(v)}[\bar{g}'(v) - I_n'(v)].$$

Differentiating with respect to v again and dividing by $H(v)$ we have

$$(17) \quad [(n-1)g(v) - n\bar{g}(v) + nI_n(v)]\frac{h(v)}{H(v)}$$

$$= \left[n + 2 - \frac{h'(v)H(v)}{h^2(v)}\right][\bar{g}'(v) - I_n'(v)] + [\bar{g}''(v) - I_n''(v)]\frac{H(v)}{h(v)}.$$

Observing that

$$(18) \qquad nI_n(v) = g(v) - I_n'(v)v \qquad \text{and} \qquad \bar{g}(v) = g(v) - \bar{g}'(v)v,$$

after some algebra, one can write (17) as

$$[n\bar{g}'(v) - I_n'(v)]\frac{h(v)v}{H(v)}$$

$$(19) \quad = \quad \left[n + 2 - \frac{h'(v)H(v)}{h^2(v)}\right][\bar{g}'(v) - I_n'(v)] + [(n+1)I_n'(v) - 2\bar{g}'(v)]\frac{H(v)}{h(v)v}.$$

Let us make the change of variables

$$(20) \qquad\qquad\qquad\qquad w(v) = \frac{h(v)}{H(v)}v.$$

Note that $w(v) \equiv 1$ corresponds to the exponential c.d.f. (1). For simplicity, we write w for $w(v)$. After multiplying both sides of (19) by $h(v)v/H(v)$, using (20) and

$$\frac{h'(v)v}{h(v)} = \frac{w'}{w}v + w - 1,$$

it is not difficult to see that (19) is equivalent to

$$(21) \qquad [n\bar{g}'(v) - I_n'(v)](w^2 - 1) = \left[(n+1)(w-1) - \frac{w'}{w}v\right][\bar{g}'(v) - I_n'(v)].$$

Clearly, $w(v) \equiv 1$ is one solution, which corresponds to the exponential distribution. It remains to prove that this is the only solution. Suppose $w(v)$ is a

solution and there exists a value $v_1 > 0$ such that $w(v_1) \neq 1$. We want to reach a contradiction. Since F is twice differentiable, we have that $w(v)$ is continuous with respect to v and hence $w(v) \neq 1$ for v in an open interval around v_1. Let $v_0 = \inf\{v | w(v) \neq 1\}$. Since $w(0+) = 1$, we have $v_0 \geq l_F$. We shall prove that $v_0 = l_F$. Assume on contrary that $v_0 > 0$. Then $w(v) = 1$ if $0 < v \leq v_0$ and integration of (20) implies that $h(v)$ is constant-valued in this interval. This contradicts the assumption (ii). Therefore $v_0 = 0$ and hence equation (21) holds for all $v > 0$. Dividing (21) by $w - 1 \neq 0$, we obtain

$$(22) \quad (w+1)[n\bar{g}'(v) - I_n'(v)] = \left[n + 1 - \frac{w'v}{w(w-1)} \right] [\bar{g}'(v) - I_n'(v)].$$

It can be seen that

$$\lim_{v \to 0+} \frac{w'v}{w(w-1)} = 1.$$

Also, using the L'Hopital's rule, one can find that

$$\lim_{v \to 0+} \bar{g}'(v) = \frac{g'(0+)}{2} \qquad \text{and} \qquad \lim_{v \to 0+} I_n'(v) = \frac{g'(0+)}{n+1}.$$

Therefore, passing to the limit in (22) as $v \to 0+$, we obtain

$$2 \left(\frac{n}{2} - \frac{1}{n+1} \right) g'(0+) = n \left(\frac{1}{2} - \frac{1}{n+1} \right) g'(0+),$$

which is not possible for $n \geq 2$. This contradiction completes the proof.

    <u>Necessity</u>. This follows from the lemma setting $r = 2$.

REFERENCES

[1] M. AHSANULLAH. Record Values – Theory and Applications. Univ. Press of America, Lanham, MD, 2004.

[2] I. S. AKHUNDOV, V. B. NEVZOROV. Conditional distributions of record values and characterizations of distributions. *J. Math. Sci.*, **145** (2007), No 2, 4837–4842.

[3] B. C. ARNOLD, N. BALAKRISHNAN, H. N. NAGARAJA. Records. Willey & Sons, New York, NW, 1998.

[4] I. Bairamov, M. Ahsanullah, A. Pakes. A characterization of continuous distributions via regression on pairs of record values. *Austr. N. Z. J. Statist.*, **47** (2005), 543–547.

[5] V. B. Nevzorov. Records: Mathematical Theory, AMS, Providence, RI, 2001.

[6] G. P. Yanev, M. Ahsanullah, M. I. Beg. Characterizations of probability distributions via bivariate regression of record values. *Metrika*, **68** (2008), 51–64.

George P. Yanev
Department of Mathematics
University of Texas – Pan American
Edinburg, Texas 78539
e-mail: `yanevgp@utpa.edu`