

Stability for the Arnold–Ghosh characterization of the geometric distribution

Jacek Wesolowski *

Mathematical Institute, Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, Poland

Received 5 November 1993; revised 17 June 1995

Abstract

Stability of the characterization of the geometric law by equidistribution of the spacing of two i.i.d. r.v.'s and one of them is studied.

AMS classification: 62E10

Keywords: Geometric distribution; Characterization of probability laws; Stability

1. Introduction

The first step in dealing with a characterization question is just to prove a characterization theorem, i.e. a result of the following type: if a distribution has some property (it fulfils a characterizing condition) then it has such and such form. Usually it is substantially less laborious than the second step: investigation of stability of the characterization. In this case we are interested in the question; if “slight” disturbances in the characterizing condition produce distributions “slightly” different from those obtained in the characterization theorem. Let us point out that such problems are of great importance for practitioners: Building a statistical model we should rely on some characteristic property of a distribution we are going to use. However, without some knowledge on stability of the characterization it may be quite risky.

Essentially researches in this area have begun in the early seventies. There is a huge literature on stability of characterizations till now (no single result for the geometric distribution is known to the author). Many of the results are gathered in a recent book by Yanushkievichius (1991), see also Kagan et al. (1973, Ch.9).

In this paper we are interested in the characterization of the geometric law given in Arnold and Ghosh (1976). It is a result making use of identical distribution of the spacing $\max\{X, Y\} - \min\{X, Y\}$ and X , where X, Y are i.i.d. natural valued r.v.'s. At

* E-mail: wesolo@alpha.im.pw.edu.pl.

first we give a short proof of the original result in Section 2. In Section 3, stability of the characterization is studied.

2. Characterization

It is well known – see Ahsanullah (1975) – that for continuous distributions equidistribution of a normalized spacing and the parent r.v. is a characteristic property of the exponential law, provided some additional technical assumptions are fulfilled. In Arnold and Ghosh (1976) a version of such result for the geometric distribution was obtained:

Theorem 1. *Let X, Y be i.i.d. natural valued r.v.'s. If for each $k = 1, 2, \dots$*

$$P(|X - Y| = k \mid |X - Y| > 0) = P(X = k),$$

then X has the geometric distribution.

Then making use of the Shanbhag lemma (see Shanbhag, 1977) the result was extended in Arnold (1980). The main aim of this paper is to consider the question of stability of the above theorem. Before that we give a short proof of Theorem 1 (being a modification of the original proof – see also Azlarov and Volodin (1986, Section 8)). A new idea of introducing the sequence (a_n) given in this proof is then fruitfully exploited in proving the stability result.

Proof of Theorem 1. From the assumption we have

$$p_r = c \sum_{k=1}^{\infty} p_k p_{r+k}, \quad r = 1, 2, \dots,$$

where $p_r = P(X = r)$, $r = 1, 2, \dots$, and $c \geq 2$ is a constant.

Define $a_1 = c p_1 = a > 0$ and $a_n = a + a_{n-1}^{-1}$, $n = 2, 3, \dots$. Then for any $n = 2, 3, \dots$ the following two implications hold:

(i) If $p_r \geq a_{n-1} p_{r+1}$, $r = 1, 2, \dots$, then $a_n p_{r+1} \geq p_r$, $r = 1, 2, \dots$, by

$$p_r = a p_{r+1} + c \sum_{k=2}^{\infty} p_k p_{r+k} \leq a p_{r+1} + a_{n-1}^{-1} c \sum_{k=2}^{\infty} p_{k-1} p_{r+k} = (a + a_{n-1}^{-1}) p_{r+1}.$$

(ii) If $a_{n-1} p_{r+1} \geq p_r$, $r = 1, 2, \dots$, then $p_r \geq a_n p_{r+1}$, $r = 1, 2, \dots$, by

$$p_r \geq a p_{r+1} + a_{n-1}^{-1} c \sum_{k=2}^{\infty} p_{k-1} p_{r+k} = (a + a_{n-1}^{-1}) p_{r+1}.$$

Now from (i), (ii) and the obvious inequality $p_r \geq a p_{r+1}$, $r = 1, 2, \dots$ we get for all $r, n = 1, 2, \dots$

$$a_{2n} p_{r+1} \geq p_r \geq a_{2n-1} p_{r+1}.$$

Hence, taking limits for $n \rightarrow \infty$, we find that $p_{r+1} = \gamma p_r$, $r = 1, 2, \dots$ for $\gamma^{-1} = \lim_{n \rightarrow \infty} a_n = (a + \sqrt{a^2 + 4})/2$. \square

3. Stability

There are different notions of stability related to types of disturbances involved in the characterization condition and in the resulting distribution. Here we are concerned with an $\varepsilon P(X = r)$ bound for the characterization condition at each point r and as a consequence a uniform estimate is obtained. It is the main result of the paper. As a corollary we obtain a bound for the total variation metric.

Denote $p_r = P(X = r)$, $r = 1, 2, \dots$ for a natural valued r.v. X .

Theorem 2. *Assume that for some $\varepsilon \in (0, 1)$ and each $r = 1, 2, \dots$*

$$|P(|X - Y| = r \mid |X - Y| > 0) - p_r| < \varepsilon p_r.$$

Then for some $\gamma \in (0, 1)$ and any $r = 1, 2, \dots$

$$|p_r - \gamma^{r-1}(1 - \gamma)| < \frac{\varepsilon}{1 - \varepsilon} K$$

where

$$K = \frac{5}{p_1^2(1 - p_1)}.$$

Proof. From the assumption we have

$$\frac{c}{1 + \varepsilon} \sum_{k=1}^{\infty} p_k p_{r+k} \leq p_r \leq \frac{c}{1 - \varepsilon} \sum_{k=1}^{\infty} p_k p_{r+k}, \tag{1}$$

where

$$c = \frac{2}{P(X \neq Y)}.$$

Now define

$$a = c p_1 > 0, \quad a_1 = a_\varepsilon = \frac{a}{1 + \varepsilon}, \quad b_\varepsilon = \frac{a}{1 - \varepsilon},$$

$$a_{2n} = b_\varepsilon + \frac{1 + \varepsilon}{1 - \varepsilon} a_{2n-1}^{-1}, \quad a_{2n+1} = a_\varepsilon + \frac{1 - \varepsilon}{1 + \varepsilon} a_{2n}^{-1}, \quad n = 1, 2, \dots$$

Step 1: In this step we will show for any $r, n = 1, 2, \dots$ the inequalities:

$$a_{2n-1} p_{r+1} \leq p_r \leq a_{2n} p_{r+1}. \tag{2}$$

The argumentation in this step is similar to that used in the proof of Theorem 1. We will prove the following two implications for any $n = 1, 2, \dots$:

- (*) If $p_r \geq a_{2n-1} p_{r+1}$, $r = 1, 2, \dots$, then $p_r \leq a_{2n} p_{r+1}$, $r = 1, 2, \dots$.
- (**) If $p_r \leq a_{2n} p_{r+1}$, $r = 1, 2, \dots$, then $p_r \geq a_{2n+1} p_{r+1}$, $r = 1, 2, \dots$.

To prove (*) observe that, by (1),

$$\begin{aligned}
 p_r &\leq b_\varepsilon p_{r+1} + \frac{c}{1-\varepsilon} \sum_{k=2}^{\infty} p_k p_{r+k} \leq b_\varepsilon p_{r+1} + \frac{c}{1-\varepsilon} a_{2n-1}^{-1} \sum_{k=2}^{\infty} p_{k-1} p_{r+k} \\
 &= b_\varepsilon p_{r+1} + \frac{1+\varepsilon}{1-\varepsilon} a_{2n-1}^{-1} \frac{c}{1+\varepsilon} \sum_{k=1}^{\infty} p_k p_{r+1+k} \leq \left(b_\varepsilon + \frac{1+\varepsilon}{1-\varepsilon} a_{2n-1}^{-1} \right) p_{r+1}
 \end{aligned}$$

On the other hand, (**) follows by (1) since

$$\begin{aligned}
 p_r &\geq a_\varepsilon p_{r+1} + \frac{c}{1+\varepsilon} \sum_{k=2}^{\infty} p_k p_{r+k} \geq a_\varepsilon p_{r+1} + \frac{c}{1+\varepsilon} a_{2n}^{-1} \sum_{k=2}^{\infty} p_{k-1} p_{r+k} \\
 &= a_\varepsilon p_{r+1} + \frac{1-\varepsilon}{1+\varepsilon} a_{2n}^{-1} \frac{c}{1-\varepsilon} \sum_{k=1}^{\infty} p_k p_{r+1+k} \geq \left(a_\varepsilon + \frac{1-\varepsilon}{1+\varepsilon} a_{2n}^{-1} \right) p_{r+1}.
 \end{aligned}$$

Now from (*), (**) and the obvious inequality $p_r \geq a_1 p_{r+1}$, $r = 1, 2, \dots$ we get (2).

Step 2: Define $b_n = a_{2n-1}$, $c_n = a_{2n}$, $n = 1, 2, \dots$. Hence (2) takes the form

$$b_n p_{r+1} \leq p_r \leq c_n p_{r+1} \tag{3}$$

for any $r, n = 1, 2, \dots$. In this step we will prove that the sequences (b_n) and (c_n) are convergent and compute their limits.

Observe that

$$c_n - c_{n-1} = \frac{1+\varepsilon}{1-\varepsilon} \frac{b_{n-1} - b_n}{b_n b_{n-1}}, \quad n = 2, 3, \dots, \tag{4}$$

$$b_n - b_{n-1} = \frac{1-\varepsilon}{1+\varepsilon} \frac{c_{n-2} - c_{n-1}}{c_{n-1} c_{n-2}}, \quad n = 3, 4, \dots$$

Hence

$$b_n - b_{n-1} = \frac{b_{n-1} - b_{n-2}}{c_{n-1} c_{n-2} b_{n-1} b_{n-2}}, \quad n = 3, 4, \dots \tag{5}$$

Since $b_2 - b_1 = ((1-\varepsilon)/(1+\varepsilon))a_2^{-1} > 0$ then by (5) we conclude that the sequence (b_n) is increasing. Hence from (4), (c_n) is a decreasing sequence. By (3), they are convergent. Denote their limits by B and C , respectively. Taking limits in (3) we obtain

$$B p_{r+1} \leq p_r \leq C p_{r+1} \tag{6}$$

for any $r = 1, 2, \dots$.

Now we find the values of B and C . The definitions of the sequences (b_n) and (c_n) yield

$$B = a_\varepsilon + \frac{1-\varepsilon}{1+\varepsilon} C^{-1}, \quad C = b_\varepsilon + \frac{1+\varepsilon}{1-\varepsilon} B^{-1}.$$

Hence

$$B = x \left(\frac{1}{1+\varepsilon} \left(a - \frac{4\varepsilon}{a} \right) \right) \quad \text{and} \quad C = x \left(\frac{1}{1-\varepsilon} \left(a + \frac{4\varepsilon}{a} \right) \right),$$

where $x(\alpha)$ is a positive solution of the equation $x^2 - \alpha x - 1 = 0$, i.e. $x(\alpha) = (\alpha + \sqrt{\alpha^2 + 4})/2$.

Step 3: In this step we prove the inequality

$$|\gamma p_r - p_{r+1}| \leq \frac{\varepsilon}{1 - \varepsilon} L p_r \tag{7}$$

for any $r = 1, 2, \dots$, where $\gamma = x^{-1}(a)$ and $L = a + 4/a$.

First observe that for any $v \geq u$

$$x^{-1}(u) - x^{-1}(v) \leq v - u \tag{8}$$

since

$$x^{-1}(u) - x^{-1}(v) = \frac{1}{2}(v - u) \left(1 - \frac{u + v}{\sqrt{u^2 + 4} + \sqrt{v^2 + 4}} \right)$$

and $\frac{|u|+|v|}{\sqrt{u^2+4}+\sqrt{v^2+4}} \leq 1$. Hence by (8) and (6)

$$\gamma p_r - p_{r+1} \leq \left[x^{-1}(a) - x^{-1} \left(\frac{1}{1 - \varepsilon} \left(a + \frac{4\varepsilon}{a} \right) \right) \right] p_r \leq \frac{\varepsilon}{1 - \varepsilon} L p_r$$

and

$$\begin{aligned} \gamma p_r - p_{r+1} &\geq \left[x^{-1}(a) - x^{-1} \left(\frac{1}{1 + \varepsilon} \left(a - \frac{4\varepsilon}{a} \right) \right) \right] p_r \\ &\geq - \frac{\varepsilon}{1 + \varepsilon} L p_r \geq - \frac{\varepsilon}{1 - \varepsilon} L p_r \end{aligned}$$

Step 4: In this step we show how the theorem follows from (7). Since

$$1 = \sum_{r=1}^{\infty} p_r = p_1 + \sum_{r=2}^{\infty} p_r = p_1 + \sum_{r=1}^{\infty} (p_{r+1} - \gamma p_r) + \gamma,$$

then by (6)

$$|p_1 - (1 - \gamma)| \leq \sum_{r=1}^{\infty} |p_{r+1} - \gamma p_r| \leq \frac{\varepsilon}{1 - \varepsilon} L \leq \frac{\varepsilon}{1 - \varepsilon} K_1,$$

where $K_1 = (a^2 + 4)(a + 1)a^{-2}$. Now for any $r = 1, 2, \dots$

$$\begin{aligned} &|p_{r+1} - \gamma^r(1 - \gamma)| \\ &\leq |p_{r+1} - \gamma p_r| + \gamma |p_r - \gamma p_{r-1}| + \dots + \gamma^{r-1} |p_2 - \gamma p_1| \\ &\quad + \gamma^r |p_1 - (1 - \gamma)| \leq \frac{\varepsilon}{1 - \varepsilon} L (p_r + \gamma p_{r-1} + \dots + \gamma^{r-1} p_1 + \gamma^r) \\ &\leq \frac{\varepsilon}{1 - \varepsilon} L \frac{1}{1 - \gamma} \\ &\leq \frac{\varepsilon}{1 - \varepsilon} K_1 \end{aligned}$$

since $1/(1 - \gamma) \leq 1 + 1/a$.

Step 5: Now the bounds for c in terms of p_1 will be found and upon applying them to K_1 the proof will be concluded. Observe that

$$c = \frac{2}{1 - P(X = Y)} = \frac{2}{1 - \sum_{k=1}^{\infty} p_k^2} \geq \frac{2}{1 - p_1^2}.$$

On the other hand

$$c = \frac{2}{P(X \neq Y)} = \frac{2}{2 \sum_{1 \leq j < k \leq \infty} p_j p_k} \leq \frac{1}{p_1 \sum_{k=2}^{\infty} p_k} = \frac{1}{p_1(1 - p_1)}.$$

Consequently

$$\frac{2p_1}{1 - p_1^2} \leq a \leq \frac{1}{1 - p_1}.$$

Now from the definition of K_1 we have

$$K_1 = \frac{[1 + 4(1 - p_1)^2](2 - p_1)(1 + p_1^2)}{4p_1^2(1 - p_1)} \leq \frac{5}{p_1^2(1 - p_1)} = K. \quad \square \tag{9}$$

Observe that the constant K_1 given on the left-hand side of inequality (9), though somewhat cumbersome, is even better than K given in the statement of Theorem 2. It should be pointed out that K and K_1 depend on p_1 in such a way that with p_1 tending to 0 or 1 the stability conditions are worsening. Question about a universal constant K remains open.

As a consequence of Theorem 2 we obtain a stability result involving the total variation metric. Recall that $\rho(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$, where μ and ν are probability measures on \mathbb{R} and \mathcal{F} denotes the Borel σ -algebra on the real line, defines a metric in the set of probability measures on \mathbb{R} . It is called the total variation metric. Denote now by μ the measure generated by the natural valued r.v. X ($P(X = r) = p_r$, $r = 1, 2, \dots$) and by ν the geometric measure with the parameter γ (defined in the Step 3 of the proof of Theorem 2).

Theorem 3. For any $\delta > 0$ there is $\varepsilon = \varepsilon(\delta) > 0$ such that if for each $r = 1, 2, \dots$

$$|P(|X - Y| = r \mid |X - Y| > 0) - p_r| < \varepsilon p_r$$

then

$$\rho(\mu, \nu) < \delta.$$

Proof. Take a natural number N such that $\mu([N, \infty)) + \nu([N, \infty)) < \tau < \delta/2$. Then for any $A \in \mathcal{F}$

$$\begin{aligned} |\mu(A) - \nu(A)| &\leq |\mu(A \cap [1, N)) - \nu(A \cap [1, N))| + \tau \\ &\leq \sum_{k=1}^{N-1} |p_k - \gamma^{k-1}(1 - \gamma)| + \tau < \frac{\varepsilon}{1 - \varepsilon} K(N - 1) + \tau \end{aligned}$$

by Theorem 2. Hence it suffices to take $\varepsilon = \frac{\tau}{K(N-1) + \tau}$. \square

Acknowledgements

Due to the referee's remark on the role of the constant appearing in the stability theorem the results of the paper were considerably improved.

References

- Ahsanullah, M. (1975). A characterization of the exponential distribution. In: *Statistical Distributions in Scientific Work*, Vol. 3. Reidel, Dordrecht, 131–135.
- Arnold, B.C. (1980). Two characterizations of the geometric distribution. *J. Appl. Probab.* **17**, 570–573.
- Arnold, B.C. and M. Ghosh (1976). A characterization of geometric distribution by distributional properties of order statistics. *Scand. Actuar. J.*, 232–234.
- Azlarov, T.A. and N.A. Volodin (1986). *Characterization Problems Associated with the Exponential Distribution*, Springer, Berlin.
- Kagan, A.M., Yu.V. Linnik and C.R. Rao (1973). *Characterization Problems of Mathematical Statistics*. Wiley, New York.
- Shanbhag, D.N. (1977). An extension of the Rao–Rubin characterization of the Poisson distribution. *J. Appl. Probab.* **14**, 640–646.
- Yanushkevichius, R. (1991). *Stability for Characterizations of Distributions*. Mosklas, Vilnius (in Russian).