On equality in distribution of some ratios involving the sum of components of a random vector

Manish C. Bhattacharjee & Sunil K. Dhar *

Center for Applied Mathematics & Statistics

Department of Mathematical Sciences

New Jersey Institute of Technology, Newark, NJ 07102

Abstract

Motivated by a classical result in the i.i.d. case for a pair random variables X, Y; we look for a simple sufficient condition, allowing for possible dependence between X and Y, under which the ratios of the components X, Y to their sum are equal in distribution. Our finding is easily extended to random vectors of higher ($n \geq 2$) dimensions to show that *exchangability* of a finite sequence $X_1, \cdots, X_n$ is sufficient to guarantee the desired result. Any *Archimedian copula* can be used as a generator of such random vectors. Our main result is applicable in many Bayesian contexts, where the observations are conditionally i.i.d. given an environmental variable with a prior.

MSC 2010 *subject classifications* Primary, Secondary, 62A01, 62N05

Keywords and phrases : random vectors, equality in distribution, exchangable finite sequences, copulas, Bayesian framework.

*Corresponding author

1 Introduction and summary

One often comes accross problems where it is questioned whether the probability of male compared to that of female are each equal to fifty percent. This question can be thought of in terms of the human sex ratio of $X : Y$ (CIA Fact Book, 2013) and the corresponding proportions being same to that of their corresponding distributions being identical. In this context X and Y are thought to be non-negative random variables. However, if the random variables X and Y are i.i.d.; it is well known that the ratios $X/(X+Y)$ and $Y/(X+Y)$ are equal in distribution. This prompts the question: if we remove the assumption of mutual independence of X and Y, can the equi-distribution of these ratios still hold, and under what reasonable conditions? In what follows, we explore some general answers to this question. We show that, if X and Y have the same distribution then $\frac{X}{X+Y}$ need not have the same distribution as $\frac{Y}{X+Y}$ and identify sufficient conditions for an affirmative answer. Extension of our main result to the case of n-dimensional random vectors $(X_1, \cdots, X_n)$ for $n \geq 2$ is indicated.

Generically, the c.d.f. of a random vector (X, Y) is denoted by $F_{X,Y}(x, y)$ and its p.d.f., when it exists, by $f_{X,Y}(x, y)$. For higher dimensional random vectors $(X_1, \cdots, X_n)$, $n \geq 2$; $F_{X_1,\cdots,X_n}(x_1, \cdots, x_n)$ and $f_{X_1,\cdots,X_n}(x_1, \cdots, x_n)$ correspondingly denote its c.d.f. and p.d.f. respectively. We use $\stackrel{d}{=}$ to denote equality in distribution of random variables (r.v.s)

2 Counterexample

We show a *counterexample* to demonstrate that $X \stackrel{d}{=} Y$ does not guarantee equality of distribution of the ratios $\frac{X}{X+Y}$ and $\frac{Y}{X+Y}$. For this purpose, we use a suitable joint density of (X, Y), that we construct via the standard normal density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{x^2}{2}\big), \qquad -\infty < x < \infty.$$

Consider the joint density function on $R^2 = (-\infty, \infty) \times (-\infty, \infty)$, given by

$$f_{X,Y}(x, y) = \big[1 + xy\phi(x)\phi^2(y)\big]\phi(x)\phi(y).$$

To see that $f_{X,Y}(x, y)$ is a valid joint density we need to observe that $\phi(x) < 1$ and that $|x\phi(x)| < 1$ (because $\frac{x^2}{2\pi} < exp(x^2)$). This in turn gives $1 + xy\phi(x)\phi^2(y) > 0$ and the fact that the mean of a scaled standard normal random variable is zero, which make $f(x, y)$ a valid density and both the marginals to be standard normal. Hence X and Y have the same distribution. We will now derive the density of $V = \frac{Y}{X+Y}$ and then show that densities of V and $1 - V = \frac{X}{X+Y}$ are not the same. Let $W = X$ note that $Y = \frac{VW}{1-V}$. The absolute value of the Jacobian is given by $\frac{|w|}{(1-v)^2}$. Hence, the joint density $f_{W,V}(w, v)$ of (W, V) on the R^2 plane is given by,

$$f_{W,V}(w, v) = f_{X,Y}\left(w, \frac{wv}{(1-v)}\right)\frac{|w|}{(1-v)^2},$$

which simplifies to,

$$\frac{|w|}{2\pi(1-v)^2}\left[1+\left(\frac{w^2v}{1-v}\right)\frac{\exp\left(\frac{-w^2}{2} - \frac{w^2v^2}{(1-v)^2}\right)}{(2\pi)^{3/2}}\right]\exp\left(\frac{-w^2}{2}\right)\exp\left(-\frac{w^2v^2}{2(1-v)^2}\right).$$

In the above joint density we integrate out the w variable, to get the marginal density of V. Note that a closed form of the density of V can be obtained by using the facts that if N is a normal random variable with mean zero and variance σ_N^2 then $E|N| = \sqrt{\frac{2}{\pi}}\sigma_N$ and $E|N|^3 = 2\sqrt{\frac{2}{\pi}}\sigma_N^3$. Hence, the density of V is given by

$$f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u, v)du = \frac{1}{\pi\big(v^2 + (1-v)^2\big)} + \frac{v(1-v)}{\sqrt{2}\pi^{5/2}\big(2(1-v)^2 + 3v^2\big)^2}.$$

Clearly, $f_V(v) \neq f_V(1-v)$, and the latter is the density of $U := \frac{X}{X+Y}$. The two ratios U and V are not equal in distribution.

Dependence between X and Y in the counterexample does not establish the necessity of their statistical independence for the equality in distribution of the ratios U, V to hold. In fact, our results are typically based on the assumption of a joint distribution, and covers independence as a special case.

3 Main results

For a random vector (X, Y), denote the ratios of the two component r.v.s to their sum, by

$$U := \frac{X}{X+Y}, \qquad V := \frac{Y}{X+Y}. \tag{3.1}$$

It may be noted that while $U + V = 1$, the r.v.s U and V cannot be thought of as the proportional contribution of the components of (X, Y) to their sum, as is obvious from the preceding counterexample.

If X, Y are absolutely continuous with a (joint) density, then so are U and V, with their respective densities related via

$$f_V(v) = f_U(1 - v). \tag{3.2}$$

Standard calculations yield an expression for the density of U. In particular, choosing the transformation

$$U = \frac{X}{X+Y}, \qquad T = X + Y\ ;$$

the joint density of (U, T) is easily seen to be $f_{U,T}(u, t) = f_{X,Y}\Big(ut\,,\,(1-u)t\Big)\,|t|$, so that the marginal density of U is

$$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}\Big(ut\,,\,(1-u)t\Big)\,|t|\,dt, \tag{3.3}$$

which together with (3.2) implies

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{\infty} f_{X,Y}\Big((1-v)t\,,\,vt\Big)\,|t|\,dt \\ &\neq \int_{-\infty}^{\infty} f_{X,Y}\Big(vt\,,\,(1-v)t\Big)\,|t|\,dt = f_U(v), \quad -\infty < v < \infty, \end{aligned}$$

in general.

Define $H(x, y)$ is *symmetric* in (x, y), if

$$H(x, y) = H(y, x), \text{ all } (x, y).$$

If however $f_{X,Y}$ has this symmetry then the above equality obviously holds. We thus have the following proposition.

Proposition 3.1 *If (X, Y) admits a joint density that is symmetric in its arguments, then the ratios in* (3.1) *are equal in distribution ($U \stackrel{d}{=} V$).*

Remark 1. There is no explicit assumption that $X \stackrel{d}{=} Y$ in the premise of the above proposition, as it is an easy consequence of the symmetry; viz,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_{-\infty}^{\infty} f_{X,Y}(y, x)\, dy = f_Y(x).$$

Remark 2. In virtue of the Remark 1 above; in the absolutely continuous case, the classic result that X, Y i.i.d. implies $U \stackrel{d}{=} V$ follows as a special case of proposition 3.1, since if X, Y are i.i.d. with a common p.d.f. $f_X(\cdot) \equiv f_Y(\cdot)$, then the joint p.d.f. satisfies

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = f_Y(x) f_X(y) = f_{X,Y}(y, x).$$

While proposition 3.1 provides an answer to our question when X, Y are absolutely continuous; an affirmative answer in the general case, where the joint c.d.f. of X, Y may also have discrete or/and singular components, is given by our next proposition.

Proposition 3.2 *If the joint c.d.f. $F_{X,Y}(x, y)$ is symmetric in (x, y), then $U \stackrel{d}{=} V$.*

Proof. With $F_{X,Y}(x, y)$ also denoting the Lebesgue-Stieltje's measure on the plane induced by the joint c.d.f., we have,

$$\begin{aligned} E(e^{itU}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(it\big(\frac{x}{x+y}\big)\right) dF_{X,Y}(x, y) \qquad (3.4) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(it\big(1 - \frac{y}{x+y}\big)\right) dF_{X,Y}(y, x) \\ &= E(e^{it(1-U)} \equiv E(e^{itV}), \qquad -\infty < t < \infty, \end{aligned}$$

where the second equality uses the symmetry condition of the joint c.d.f.. Thus, the ratios U and V having the same *characteristic function* must be equal in distribution.

The joint c.d.f.'s symmetry condition was motivated by the corresponding assumption in Proposition 3.1 and the following observation.

Lemma 3.3 (i) *Suppose X, Y are absolutely continuous. Then $F_{X,Y}(x,y)$ is symmetric in its arguments (x,y) if and only if so is $f_{X,Y}(x,y)$.*

(ii) *The symmetry condition in Proposition 3.2 implies X and Y are identically distributed.*

Proof. (i) Suppose $f_{X,Y}(x,y)$ is symmetric in (x,y). Then the non-negativity of the integrand and Fubini's theorem implies,

$$\begin{aligned} F_{X,Y}(x,y) = P(X \leq x,\ Y \leq y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\ dv\ du \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(v,u)\ dv\ du \\ &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(v,u)\ du\ dv \\ &= P(X \leq y,\ Y \leq x) \equiv F_{X,Y}(y,x). \end{aligned}$$

Conversely, supposing $F_{X,Y}$ is symmetric in its argument (x,y), and has a joint density; we have,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x, \partial y} F_{X,Y}(x,y) = \frac{\partial^2}{\partial x, \partial y} F_{X,Y}(y,x) = f_{X,Y}(y,x).$$

(ii) Using the pointwise symmetry of $F_{X,Y}(\cdot,\cdot)$ on R^2,

$$P(X \leq x) = \lim_{y\to\infty} F_{X,Y}(x,y) = \lim_{y\to\infty} F_{X,Y}(y,x) = P(Y \leq x).$$

Remark 3. The symmetry condition in Proposition 3.2 is of course equivalent to X, Y being '*exchangable*', i.e., $(X,Y) \stackrel{d}{=} (Y,X)$. For a pair of random variables however, it is much more simply stated as the property that the joint c.d.f. $F_{X,Y}(\cdot,\cdot) : R^2 \longrightarrow [0,1]$ is symmetric in its arguments. For random vectors of higher dimensions, the corresponding condition that the c.d.f. $F_{X_1,\cdots,X_n}(x_1,\cdots,x_n)$ is permutation invariant in its arguments is more succinctly and elegantly descried as $X_1,\cdots,X_n$ being exchangable; thus generalizing our earlier proposition as follows.

Proposition 3.4 *If $X_1,\cdots,X_n$ $(n \geq 2)$ is a finite exchangable sequence,*

then

$$\frac{X_j}{S_n} \stackrel{d}{=} \frac{X_k}{S_n}, \qquad j,k \in \{1,2,\cdots,n\},\ j \neq k$$

where $S_n := \sum_{i=1}^n X_i$.

Proof. Suppose $X_1, \cdots, X_n$ $(n \geq 2)$ are exchangable; i.e., $(X_{i_1}, \cdots, X_{i_n}) \stackrel{d}{=} (X_1, \cdots, X_n)$ for all permutations $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$. For brevity, denote by

$$\begin{aligned} \boldsymbol{X} &:= (X_1, \cdots, X_n), \text{ and} \\ 0_j\boldsymbol{X} &:= (X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_n), \end{aligned}$$

be the corresponding vector that skips the j-th coordinate X_j, and the corresponding values assumed by, $\boldsymbol{x}$ and $0_j\boldsymbol{X}$ respectively. When

$$\begin{aligned} E\left\{\exp\left(it\frac{X_j}{S_n}\right)\right\} &= \int_{-\infty}^{\infty} \exp\left(it\frac{u}{s_n}\right) dF_{\boldsymbol{X}}(x_1, \cdots, x_{j-1}, u, x_{j+1}, \cdots, x_n) \\ &= \int_{-\infty}^{\infty} \exp\left(it\frac{u}{s_n}\right) dF_{(X_j, 0_j\boldsymbol{X})}(u, 0_j\boldsymbol{x}) \\ &= \int_{-\infty}^{\infty} \exp\left(it\frac{u}{s_n}\right) dF_{(X_k, 0_k\boldsymbol{X})}(u, 0_k\boldsymbol{x}) \\ &= E\left\{\exp\left(it\frac{X_k}{S_n}\right)\right\}, \end{aligned}$$

where the value s_n of S_n is given by $s_n = u + \sum_{i=1_{i\neq j}}^n x_i$ or $s_n = u + \sum_{i=1_{i\neq k}}^n x_i$ in the second or third integrands above, respectively. Note, the two equalities preceding the last step hold since $(X_j, 0_j\boldsymbol{X}) \stackrel{d}{=} \boldsymbol{X} \stackrel{d}{=} (X_k, 0_k\boldsymbol{X})$ for all pairs j, k, by exchangability. Hence the result.

In conclusion, any *Archimedian copula* can be used as a generator of such exchangable random variables, Nelson (1999) and Genest et al. (1986). These results are also applicable to Bayesian contexts, where the observations are conditionally i.i.d. given an environmental variable with a prior distribution.

References

[1] The Central Intelligence Agency of the United States (2013), *CIA Fact Book.* https://www.cia.gov/library/publications/the-world-factbook/index.html.

[2] Nelsen, R.B. (1999), *An Introduction to Copulas.*Lecture Notes in Statistics, # 139; Springer, New York.

[3] Genest, C., and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *Amer. Statist.* **40**, 280-285.