Some Properties of the WJ Distribution and Implication in Information Theory

Geying Liang,1,a, Han Xue,1, Qiong Jia,2,3,b and Junhua Wu1,c

1College of Digital Media & Peter Grünberg Research Center, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2Department of Management, Hohai University, Nanjing 211100, China;
3Department of Operation and Logistics, Sauder Business School, University of British Columbia, Vancouver BC V6T 1Z2, Canada.

a lianggeying@126.com (G.Y.L.); b jiaqionghit@163.com (Q.J.); c wjhtsinghua@163.com (J.H.W.).

Abstract. The WJ probability density distribution function describes a general mechanism for various stochastic processes including extreme events and critical phenomena. This work investigates the potential application of the WJ distribution in information theory, by means of exploring the distribution itself, the probability density distribution function of information entropy and an expression for relative information entropy. Changing the multiple parameters of the function, the WJ probability density distribution function as well as the corresponding information entropy function distribution and relative information entropy is systematically analysed and compared. The characteristics of the WJ probability density distribution function and information entropy function are explicitly manifested, showing application prospective of the distribution in information theory.

1. Introduction

In many models of probability distributions [1-3], each shows its distinct features for a unique purpose. For instance, the uniform distribution as a constant addresses the probability density of a random variable over a certain interval, i.e., it is equally possible to take a value within the interval [4]. The Gaussian distribution, which is widely used, indicates the concentrated occurrence of a variable in some domain, with the fast diminishing probability outside [5]. Such a probability law, though, describes a usual event occurrence quite satisfactorily, it is challenged in handling more complex phenomena such as in the case of so-called extreme events [6-14]. The WJ probability density distribution, as shown below, deliberates a more general distribution property that can be used to represent the occurrence and extremum of rare events [8]. It has merits capable of covering the scenarios of common distribution functions by appropriate changing of the parameters therein. An example is the parametric selection of the WJ distribution to virtually simulate a Gaussian distribution, but not vice versa in general.

It is vital in the era of information that people create, use, transmit and filter out useful information as an explosive number of messages is available to extract and communicate [15,16]. Such a collection, transmission and analysis process of big information data may be displayed in a statistical manner [15-18]. This study aims at numerical analyses of the WJ distribution and its implication in information theory.
2. Theoretical considerations

2.1 The WJ Distribution

The expression of the WJ probability density function of a global random variable is given by [8]

\[ f(x, \alpha, \beta, \mu) = \frac{\beta}{\Gamma(\alpha / \beta)} e^{-\frac{\beta}{\alpha}(x-\mu)^{\alpha}} e^{-\frac{\beta}{\alpha} x^\beta} \]  

(1).

The function has three parameters as \( \alpha, \beta \) and \( \mu \). The magnitudes of \( \alpha \) and \( \beta \), in determining the nature of an infinite number of small asymmetric fluctuations around the mean of a constant random variable, dominates major probabilistic properties of the global variable, particularly the shape and skew of the distribution. The average value \( \mu \) adjusts the position of the probability density distribution, but doesn’t affect the shape and symmetry of the distribution. Moreover, \( \alpha \) and \( \beta \) may have an additional contribution to the location of the distribution by shifting \( \frac{1}{\beta} \psi(\frac{\alpha}{\beta}) \). Since the change in the mean \( \mu \) horizontally transfers the distribution curve instead of the symmetry and shaping properties, it is set to zero in the paper for brevity. So, the research focuses on investigating properties of the remaining two parameters \( \alpha \) and \( \beta \) through setting a series of values of \( \alpha \) and \( \beta \), revealing the symmetry and diversity of the distribution.

2.2 Information entropy vs. probability density function

In information theory, entropy measures the uncertainty of the state of a message from a source. For the information entropy of a (continuous) source, it is useful to express it in the form of

\[ H = - \int f(x) \ln f(x) dx \]

(2)

in terms of its probability density distribution function \( f(x) \) [17,18]. Often the integral takes places over the whole domain \((-\infty, +\infty)\), resulting in the total information entropy.

In this work, a slightly different expression is defined as follows

\[ H(x) = - \int_{-\infty}^{\infty} f(x, \alpha, \beta, \mu) \ln f(x, \alpha, \beta, \mu) dx \]

(3)

as a measure of relative entropy of the WJ probability density function.

2.3 Analysis tools

This work used the software packages of Mathematica, MATLAB and Origin for analyzing the distribution functions and presenting the results.

3. Results and discussion

The research systematically changes the parameter set of \( \alpha \) and \( \beta \) to study the properties of the WJ function, the probability density distribution function of the function information entropy \( -f(x, \alpha, \beta, \mu) \ln f(x, \alpha, \beta, \mu) \) and the relative information entropy \( H(x) \). The outcomes are divided in two major classes, one with the variation of \( \alpha \) for a given \( \beta \) and the other with the variation of \( \beta \) for a given \( \alpha \).

3.1 Outcomes of varying \( \beta \) for a given \( \alpha \)

Addressed firstly are the results from a smaller parameter \( \alpha \) and a set of smaller \( \beta \). Fig. 1 shows the variation of \( \beta \) from the probability density curve of 0.1-1 for \( \alpha = 0.1 \). In Fig. 1(a) it is observable that the function has a non-standard symmetry. The larger a \( \beta \) value, the higher the peak and the sharper
the graph. Nonetheless, the overall trend shows a gradual increase from a minimum value, then to a maximum value, and finally to a mild diminution towards zero. As shown in Fig. 1(b), the corresponding information entropy density distribution curve conforms with the change trend of the probability density distribution function curve. In Fig. 1(c), the relative information entropy curve is apparently symmetric about the origin owing to the definition. The consideration is solely given to the part of the curve when \( x > 0 \). Changing with \( x \), the function change rate becomes smaller and smaller, and the final function change rate is 0, or the function is finally stable at a certain value, i.e., the total information entropy.

![Fig. 1. Cases for \( \alpha = 0.1 \) and \( \beta \) from 0.1 to 1.0 (c.f. the figure for specific parameters).](image)

Then \( \alpha \) is increased to 0.5 and the same set of \( \beta \) is retained as shown in Fig. 2. Clearly, the probability density distribution function curve becomes narrower and sharper, while the information entropy probability distribution curves changes not as significantly when compared to Fig. 1(b). The larger the \( \beta \) value, the larger the maximum probability density, indicating that the probability is confined within a certain range. So, the function curve varies greatly in a small interval. The trend of the information entropy function is similar to that of Fig. 1, and the corresponding total (maximum) information entropy turns to become smaller.

![Fig. 2. Cases for \( \alpha = 0.5 \) and \( \beta \) from 0.1 to 1.0 (c.f. the figure for specific parameters).](image)

When the value of \( \alpha \) increases from 0.5 to 1, the domain range of the curve with noteworthy values shrinks further and the function value augments. There is a symmetrical tendency with respect to \( x = 0 \), which is analogous to the feature of a Gaussian density function. The information entropy probability distribution curve relatively reaches asymptotically a plateau stable rather quickly, and possesses a similar trend to the probability density distribution function. This fact proves that when \( \alpha \) and \( \beta \) are small, the information entropy distribution is concentrated and the change is relatively small. The information entropy curve trends are consistent with that of the WJ probability density distribution function. When \( x > 0 \), the final steady information entropy maximum decreases as the \( \alpha \) value increases.
After analysing the change of probability density function and information entropy in the moderate variation of $\alpha$ from 0.1 to 1, more extreme values are explored to learn the impacts of enlarging the value of $\alpha$ and $\beta$ on the behaviour of the functions and the curves of the information entropy probability distribution.

Fig. 4 shows the phenomena from the value of $\beta$ being 1, 5, 10, 20, 50, and 100 when $\alpha = 1$. Obviously, the probability density distribution becomes larger and the curve appears more asymmetric. For the fixed $\alpha$ value, a larger $\beta$ leads to the much steeper increasing trend of the curve. Meanwhile, novel features emerge in the information entropy probability distribution curve. It is visible that the information entropy probability distribution in the figure splits into two components, accompanying a minimum value interval. The information entropy tends to become smaller, but the trend of information entropy curves stays not much different.

From the above discussion, increasing the values of $\alpha$ and $\beta$ may render the curves more and more sharp. It is true that further increasing $\alpha$ changes the performance of the distributions radically, as exposed in the curves of $-\ln f(x, \alpha, \beta, \mu)$ from positive to negative and the corresponding negative information entropy, confines the substantial contribution of the functions to a much smaller interval [19-21]. For the emerging of negative entropy, it is worthy to further study and results shall be presented elsewhere. When the domain of $x$ is magnified, it becomes clearer that the probability density function has a sharp rise and rapid decline in the interval of -1 to 1. The probability
density distribution function curve begins to exhibit a symmetric style with respect to increasing $\beta$, and the information entropy distribution continues to reduce, promoting the relative information entropy from positive to negative. In the figure, the information entropy curves when $x>0$ show that the negative entropy increases in magnitude as $\beta$ increases. In addition, the information entropy function traces a minimum value interval.

![Fig. 6. Cases for $\alpha=50$ and $\beta$ from 1 to 100 (c.f. the figure for specific parameters).](image)

![Fig. 7. Cases for $\alpha=100$ and $\beta$ from 1 to 100 (c.f. the figure for specific parameters).](image)

A great increase in the $\alpha$ value can acquire almost perfect symmetry curves for the same $\beta$, as respectively shown in Figs. 6 and 7. In specific, the information entropy probability distribution curves show a sharp downward trend. The information entropy maintains the same trend as in Figure 5, but its negative entropy value diminishes steadily. In this experiment of changing the $\alpha$ value, a rule may be deduced. The larger the $\alpha$ value, the larger the function value of the distribution and the more severe the curve shrinkage. Comparing the curves of the same $\alpha$ value, it is found that the larger the value of $\beta$, the steeper the curve. The information entropy probability distribution relatively keeps temperately stable when $\alpha$ and $\beta$ are small, and there is not much wavering. When $\alpha$ and $\beta$ are large, the negative value appears and can be large. Under the condition, the information entropy becomes negative in the range of $x>0$ and the corresponding value is getting smaller as $\alpha$ and $\beta$ increases.

### 3.2 Outcomes of varying $\alpha$ for a given $\beta$

The following sections turn to discuss the cases of varying $\alpha$ for a given $\beta$, in the way akin to the cases of varying $\beta$ for a given $\alpha$ as tackled in the precedent sections.
The curves shown in Fig. 8 are the results under the conditions of $\alpha$ varying from 0.1 to 1 for $\beta$ set to 0.1. It is straightforward that the smaller the $\alpha$ value, the less symmetrical the curve when $\beta$ is fixed at 0.1. Nevertheless, the overall trend for Fig. 8(a) is comparable in the form of first increasing to a peak value and then decreasing in turn. Comparing Figs. 1 and 8 shows that the curves of the functions have a faster falling progression from changing the $\beta$ value than changing the $\alpha$ value. Moreover, the probability maximum position has a rightward shift when the $\alpha$ value increases. The curves of the information entropy probability distribution demonstrate the same trend. The inclination in the corresponding information entropy is consistent with the change of the parameter $\alpha$. It is remarkable that the information entropy has a more sensitive dependency on $\beta$ for the sceneries in Fig. 1(c) and Fig. 8(c).

In terms of Fig. 9(a) to Fig. 10(a), the symmetry of the curve does not change much for the same $\alpha$ after increasing the $\beta$ value, in contrast to that the function peak value becomes larger yet the magnitude of the enlargement is not conspicuous. Both the information entropy probability distribution and information entropy share similar propensities for $\beta=0.5$ and $\beta=1$, as presented in Figs. 9(b, c) and 10(b, c).

Subsequently, the experiment looked at more extreme cases when $\alpha=1$ to 100 at the fixed $\beta$ of 1. The WJ distribution becomes faster concentrated around the origin as $\alpha$ increases, as shown in Fig.
11(a). In Fig. 11(b), the radical splitting of the information entropy probability distribution happens for it gives a multiple-peaked curve as $\alpha$ increase from 1 to 100. The switching of the relative information entropy from positive to negative is enormous as $\alpha$ increases, as documented in Fig. 11(c).

![Fig. 11. Cases for $\beta=1$ and $\alpha$ from 1 to 100 (c.f. the figure for specific parameters).](image)

Fig. 12 shows the results under the conditions of $\beta=10$ and $\alpha=1, 5, 10, 20, 50, \text{and} 100$, respectively. With increasing $\alpha$, it is obvious that the distribution is more spread and the function symmetry is biased, as shown in Fig. 12(a). Besides, the curve for the information entropy probability distribution unveils a state in which the negative division is large. The larger $\alpha$ is, the sharper the curve is. Parallel to $\beta=1$, the cases of $\beta=10$ for the corresponding information entropy have negative entropies and a minimum value interval ensues. The trend follows the change of the extreme value with respect to the situation of the fixed $\alpha$ in the previous discussion. When $x>0$ is observed, the final steady information entropy maximum decreases as $\alpha$ increases.

![Fig. 12. Cases for $\beta=10$ and $\alpha$ from 1 to 100 (c.f. the figure for specific parameters).](image)

As shown in Figs. 13 and 14 for $\beta=50$ and $\beta=100$ separately, the corresponding behaviour of the WJ distribution, the information entropy probability function and the relative information entropy is alike that of $\beta=10$. Alternatively speaking, it is clearly observable based on the $\beta$ value from 10 to 100 that the larger the value of $\alpha$ is, the larger the value of the distribution, as well as that the more obvious the shrinking effect is, the more obvious the symmetry is. Nonetheless, when the $\beta$ value is
larger than the $\alpha$ value by several tens of times or even hundreds of times, the distribution value is small, and there is no obvious symmetrical state, along with the observation that the distribution is relatively spread rather than concentrated. The information entropy probability distribution displays similar manners. When the $\beta$ value is greater than the $\alpha$ value, the information entropy is larger, while the distribution is more concentrated, the information entropy is small.

![Fig. 14. Cases for $\beta = 100$ and $\alpha$ from 1 to 100 (c.f. the figure for specific parameters).](image)

4. Conclusions

This work has studied the prospective inference of the WJ distribution in information theory, via deliberating the distribution itself, the probability density distribution function of information entropy and the relative information entropy. Varying the multiple parameters of the function, the WJ probability density distribution function as well as the corresponding information entropy function distribution and relative information entropy has been systematically analysed and compared. The characteristics of the WJ probability density distribution function and information entropy function are explicitly manifested, showing that the WJ probability density function has more general distribution characteristics and application potential in information theory and other relevant disciplines.

Acknowledgements

This work was in part supported by the National Natural Science Foundation of China (Nos. 31770361 & 71702045), the Fundamental Research Funds for the Central Universities (No. 2013B1802026, 2014B14414), Humanities and Social Sciences Foundation of the Ministry of Education in China (No.16YJC630028, 17YJC630047), Nano and Material Technology Development Program through the National Research Foundation of Korea (No. 2014M3A7B4052193), and by the Ministry of Trade, Industry and Energy of Korea under Industrial Technology Innovation Program (No.10080408), the Science and Technology Innovation Training Program of Jiangsu Province, China (No. 201810293063Y), and the Special Funds of Nanjing University of Posts and Telecommunications of China (NUPTSF, Grant Nos. NY215028 & NY217025).

References

[1] Freedman, D. A. (2005) Statistical Models: Theory and Practice, Cambridge University Press. ISBN 978-0-521-67105-7.
[2] Olofsson, P. (2005) Probability, Statistics, and Stochastic Processes, Wiley-Interscience. ISBN 0-471-67969-0.
[3] Parzen E. On estimation of a probability density function and mode[J]. Ann. Math. Statis. 1962, 33(3):1065-1076.
[4] Nechval, K. N., Nechval, N. A., Vasermanis, E. K., Makeev, V. Y. Constructing shortest-length confidence intervals. Transport and Telecommunication 3 (1) 95-103 (2002).
[5] Bryc, W. (1995) The Normal Distribution: Characterizations with Applications. Springer-Verlag. ISBN 0-387-97990-5.
[6] Bramwell, S. T. et al. Universal fluctuations in correlated systems. Phys. Rev. Lett. 84, 3744–3747 (2000).
[7] Weinreich, D. M., Delaney, N. F., DePristo, M. A., Hartl D. L. Darwinian evolution can follow
rare mutational paths to fitter proteins. Science 312, 111–114 (2006).
[8] Wu, J. H. and Jia, Q. A universal mechanism of extreme events and critical phenomena. Sci. Rep. 6, art. no. 21612; doi: 10.1038/srep21612 (2016).
[9] Goncalves, R., Ferreira, H. & Pinto, A. A. Universality in the stock exchange market. J. Diff. Eq. Appl. 17, 1049–1063 (2011).
[10] Alfinito, E., Reggiani, L. Evidence of Gumbel distributions of conductance fluctuations in bacteriorhodopsin thin films. J. Phys.: Condens. Matter 25, 375103, 10.1088/0953-8984/25/37/375103 (2013).
[11] Cai, W. et al. Increasing frequency of extreme El Niño events due to greenhouse warming. Nature Climate Change 4, 111–116 (2014).
[12] Varotsos, P. A., Sarlis, N. V., Tanaka, H. K., Skordas, E. S. Similarity of fluctuations in correlated systems: The case of seismicity. Phys. Rev. E 72, 041103, 10.1103/PhysRevE.72.041103 (2005).
[13] Spinney, R. E., Prokopenko, M., Lizier, J. T. Transfer entropy in continuous time, with applications to jump and neural spiking processes. Phy Rev E. 95(3-1), 032319 (2017). doi: 10.1103/PhysRevE.95.032319.
[14] Ghil, M. et al. Extreme events: dynamics, statistics and prediction. Nonlin. Processes Geophys. 18, 295–350 (2011).
[15] Gleick, J. (2011) The Information: A History, a Theory, a Flood. Random House Digital. ISBN 978-0-375-42372-7.
[16] Cover, T. M., Thomas, J. A. (2006) Elements of Information Theory. John Wiley and Sons.
[17] Waldram, J. R. (1985) The Theory of Thermodynamics. Cambridge University Press. ISBN 978-0-521-28796-8.
[18] Huang, K. (1987) Statistical Mechanics, 2ed. Random House Digital. ISBN 978-0-471-81518-1.
[19] Abramsky, S., Brandenburger, A. (2014) An Operational Interpretation of Negative Probabilities and No-Signalling Models. In: van Breugel F., Kashefi E., Palamidessi C., Rutten J. (eds) Horizons of the Mind. A Tribute to Prakash Panangaden. Lecture Notes in Computer Science, vol. 8464. Springer, Cham.
[20] Abramsky, S., Brandenburger, A.: The sheaf-theoretic structure of non-locality and contextuality. New Journal of Physics 13, 113036 (2011).
[21] Burgin, M. Interpretations of negative probabilities. arXiv preprint 1008.1287 (2010).