

Research Article

An Alternate Formulation for Computing/Validating the Shannon Entropy of Probability Distributions

Parthasarathy Srinivasan* 

Oracle Corporation, Lehi, the United States

Abstract

One of the most pervasive applications in Computing, is the generation of Random numbers, which belong to a certain probability distribution such as a Gaussian (normal) distribution. These probability distributions possess statistical properties such as expected values (mean), variance (standard deviation), p-value, Entropy etc.; out of which Entropy is significant, for quantifying the amount of (useful) information, that a particular instance of a distribution embodies. This quantification of Entropy is of value as a characterizing metric, which determines the amount of randomness/uncertainty and/or redundancy that can be achieved using a particular distribution instance. This is particularly useful for communication, cryptographic and astronomical applications in this day and age. In the present work the Author introduces an alternate way to calculate the approximate value of the Information Entropy (with a variation to the formulation of Information Entropy by Claude Shannon, as known by the scientific community); by observing that a Takens embedding of the probability distribution yields a simple measure of the Entropy; by taking into consideration only four critical/representative points of the embedding. By comparative experimentation, the Author has been able to empirically verify that this alternate formulation is consistently valid: The baseline experiment chosen relates to Discrete Task Oriented Joint Source Channel Coding (DT-JSCC) which utilizes entropy computation to perform efficient and reliable task oriented communication (transmission and reception) as will be elaborated further. The author performed the comparison by employing the Shannon formulation for Entropy computation in the baseline DT-JSCC experiment and then repeating the experiment by employing the Entropy formulation, introduced in this work. Eventually, the accuracy of results obtained (data models generated) were almost identical (differing in accuracy by only ~ 1% overall). Thus, the alternate formulation introduced in this work, provides a reliable means of validating the random numbers obtained from the Shannon formulation and also potentially serves as a simpler, faster, and more computationally optimal method. This is particularly useful in applications, where there is a constraint on the computational resources available, such as mobile and limited devices. The method is also useful as a way of uniquely identifying and characterizing Random probability sources, such as those from astronomical and/or optical (photonic) phenomenon. The author also investigates the impact of incorporating the above notion of Entropy into the Mars Rover IER software and confirms the conclusions in the original article from Jet Propulsion Laboratories, NASA, which describes the ICER Progressive Wavelet Image Compressor.

Keywords

Shannon Entropy, Alternate Formulation, Golomb Codes, Takens Embedding

*Corresponding author: parthasarathy.srinivasan@oracle.com (Parthasarathy Srinivasan)

Received: 10 November 2025; **Accepted:** 21 November 2025; **Published:** 24 December 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

In the DT-JSCC approach the source dataset is “discretely encoded” by machine learning techniques [11-13] which utilize a discrete codebook of symbols to comparatively extract and represent task-oriented semantic information/data from the input data source.

This extracted information is processed to introduce redundancy in proportion to the computed entropy of the signal, in order to ensure robustness of the communicated messages against tampering/distortion during transmission and reception.

Turns out that the entropy computation using traditional Shannon formulation [9] and that computed using the formulation introduced in this work yield almost identical accuracy in the evaluation of the recognition/classification task w.r.t. the data that is input at the source end and that which is received at receiver end.

This lends (unequivocal) credence to the validity of the Entropy formulation introduced in this work, as an alternative means to compute/validate Shannon Entropy.

The additional experiment with the Mars Rover IER software confirms the findings that the amount of redundancy that the IER software (in its original form) introduces/utilizes is indeed optimal.

In the IER experiment [4] the input image is raster scanned and a notion of “Probability of Zero” is utilized for lossless image compression. The raster scanned image is segmented into well-defined zones or segments and pixels are processed in accordance with the zones in which the pixel falls along with the adjacent pixels.

The IER software keeps a running count of pixels estimated to be zero pixels and uses this to compute a moving probability statistic with each additional pixel being encoded in line with this probability.

The decoder component of the IER software inverts the above transformation and reconstructs the output image from the encoder output. One of the key questions investigated by the IER team relates to the optimality of the above process w.r.t. the amount of redundancy of information present in the compressed image form under varying parameters such as the choice of designation for a bit to be processed as a zero bit or vice versa. The IER team measures and plots the compression ratio against the designated zero bit probabilities and finds the scheme in use to be optimal and concludes that any changes in this aspect yield suboptimal results only.

The author of this present work explored the possibility of changing the above aspect by modifying the scheme to designate the Probability of Zero to a bit by computing and utilizing moving entropy information. The results of the above modified experiment lead to the same conclusion that the IER scheme is already optimal.

This indicates that the notion of entropy introduced by this author can also be used as an ancillary test to validate hypothesis such as the above across various experiments.

2. Literature Survey

The Author conceived this new formulation for Entropy by employing critical extrapolative reasoning to link emergent notions from the Authors own previous works. Once formulated the Author was on the lookout for experiments to be found in existing scientific literature to corroborate or disprove the validity of the formulation. In this context the Discrete Task Oriented JSSC experiment [1, 7] was chosen as a typical baseline candidate due to the simplicity yet exhaustiveness of the experiment [11] to verify and establish the validity of the formulation of Entropy introduced in this work.

The author found the DT-JSCC experiment by searching the Arxiv Data/Information Base.

Other related literature may be found by further search on the same/similar platform by traversing the reference flower provided by the platform.

As mentioned above, the author also found the IER software experiment to be schematically suitable for yet another test to incorporate and substantiate the entropy measure and its related potential applications. Particularly, the IER experiment is chronologically significant in that the original experiment has evolved over a decade and has remained pretty much complete and conclusive in light of newer methods and advances in scientific reasoning. The context modeling scheme used by ICER, utilizes previously encoded information from spatially neighboring coefficients.

The “Probability of Zero” predictions made by the ICER encoder relies on the concept of “Golomb codes”.

ICER determines segment boundaries for the LL (non-contrast enhanced) sub band first, and then maps them to the other sub bands.

The trade-off between compression volume and image quality is somewhat mitigated by the ICER approach of progressive compression (similar to JPEG [5, 6]), which can meet a stipulated compression rate constraint. The “byte-quota” measure introduced by ICER controls the above trade-off for different modes of operation.

Error containment is implemented in ICER by segmenting the image into distinct fragments so that any loss of information in a particular segment does not impact the encoding and representation of other segments.

2.1. A Note about Golomb Codes

The Golomb code is a prefix code particularly useful when smaller values occur with greater probability than larger values leading to a Geometric distribution modelling the alphabet occurrence.

Golomb coding uses a tunable parameter M to divide an input value x into two parts: q , the result of a division by M , and r , the remainder. The quotient is sent in unary coding, followed by the remainder in truncated binary encoding.

2.2. A Note About Takens Embedding [14]

The Takens embedding creates a new time series spanning multiple enumerated dimensions from a time series of single dimension, representing the evolution of the time series.

The output of the Takens Embedding process is a flattened 2D array which represents the embedded data points.

The optimal values for d (the embedding dimension) and τ (the time delay) are determined using methods like the minimum mutual information for τ and the nearest neighbors method for d .

C-(core) code snippet for takens embedding

```
// Loop to create each embedded point
for (i = 0; i < num_embedded_points; i++) {
// Loop to fill the dimensions of the current embedded point
for (j = 0; j < d; j++) {
embedded_data[i * d + j] = data[i + j * tau];
}
}
```

Comparative Experimentation

Experiment I: DT-JSCC

Simulation of the MINST dataset comprising the tasks of encoding, modulation, transmission, reception, demodulation, and decoding, was performed for the baseline case and the comparative case under evaluation with measurement of communication accuracy.

The following code reference from [1, 2] was adapted to perform the comparative experiment:

```
@article{xie2022robust,
title={Robust Information Bottleneck for Task-Oriented
Communication with Digital Modulation},
author= {Xie, Songjie and Wu, Youlong and Ma, Shuai and
Ding, Ming, and Shi, Yuanming and Tang, Mingjian,
journal={arXiv preprint arXiv:2209.10382},
year={2022}
}
```

The MNIST dataset was processed with the Classical Shannon Entropy formulation and the Variant formulation for Entropy introduced by the author in this work as detailed below (python implementation provided for reference):

```
def _entr(dist):
dist = dist + 1e-7
# The commented code represents the shannon formulation
and the functional code is the authors'(variant) formulation of
Entropy [3]
```

```
# en_z_M = torch.mul(
# -1*dist, torch.log(dist)
#)
```

```
# en_z = torch.sum(
# torch.sum(en_z_M, dim=-1),
# dim=-1)/en_z_M.size(-2)
emb = takens(dist)
x=emb[:, 0]
y=emb[:,1]
#z=[0 for i in range(len(emb[:, 1]))]
#mod_ent=np.zeros((77))
minprob = np.argmin(x.T)
maxprob = np.argmax(x.T)
minchg = np.argmin(y.T)
maxchg = np.argmax(y.T)
print(minprob,maxprob,minchg,maxchg)
print(x[minprob:,0])
print(x[maxprob:,0])
print(x[minchg:,0])
print(x[maxchg:,0])
a1=0.0001
a2=0.0001
a3=0.0001
a4=0.0001
if(np.size(x[minprob:,0])!=0):
a1=np.argmin(x[minprob:,0])
if(np.size(x[maxprob:,0])!=0):
a2=np.argmin(x[maxprob:,0])
if(np.size(x[minchg:,0])!=0):
a3=np.argmin(x[minchg:,0])
if(np.size(x[maxchg:,0])!=0):
a4=np.argmin(x[maxchg:,0])
print('***')
print(a1,a2,a3,a4)
en_z =
abs(abs(a1+0.0001)*log(abs(a1+0.0001),2)+abs(a2+0.0001)
*log(abs(a2+0.0001),2)+abs(a3+0.0001)*log(abs(a3+0.0001
),2)+abs(a4+0.0001)*log(abs(a4+0.0001),2))
#mod_ent[k] =
abs(abs(probs[k]-x[minprob]+errfact)*log(abs(probs[k]-x[mi
nprob]+errfact),2)+abs(probs[k]-x[maxprob]+errfact)*log(ab
s(probs[k]-x[maxprob]+errfact),2)+abs(probs[k]-x[minchg]+
err-
fact)*log(abs(probs[k]-x[minchg]+errfact),2)+abs(probs[k]-x
[maxchg]+errfact)*log(abs(probs[k]-x[maxchg]+errfact),2))
return en_z
```

3. Results

Experiment I

```

Administrator: Command Prompt

Directory of C:\Users\parth\Desktop\results

03/30/2025  10:53 AM  <DIR>          .
03/30/2025  11:00 AM  <DIR>          ..
03/30/2025  10:36 AM                17,820 CIFAR10-num_e16-num_latent4-modpsk-psnr4.0-lam0.001.pt
03/30/2025  10:48 AM                161 rpytfile.py
                2 File(s)            17,981 bytes
                2 Dir(s)      789,610,930,176 bytes free

C:\Users\parth\Desktop\results>python rpytfile.py
{'acc1s': [89.521484375, 90.029296875, 90.439453125, 90.859375, 90.673828125, 90.732421875, 90.654296875, 90.771484375,
90.673828125, 90.556640625], 'acc3s': [97.353515625, 97.392578125, 97.626953125, 97.890625, 97.783203125, 97.7734375, 97.
548828125, 97.685546875, 97.685546875, 97.548828125], 'dist': tensor([[9.7839e-10, 2.1176e-07, 2.8868e-07, ..., 1.0869
e-10, 1.3304e-08,
3.9888e-07],
[1.1608e-12, 6.7268e-09, 1.1593e-08, ..., 4.9211e-10, 4.5897e-09,
2.0134e-08],
[6.8405e-23, 7.5416e-16, 8.5699e-14, ..., 1.8219e-22, 2.3172e-19,
1.3962e-17],
...,
[9.9985e-01, 1.8282e-05, 3.0085e-06, ..., 1.3377e-08, 9.1097e-07,
9.3622e-06],
[1.0000e+00, 4.5008e-08, 1.0406e-09, ..., 8.7299e-08, 7.6560e-08,
2.7651e-07],
[1.0000e+00, 3.5330e-07, 9.4388e-09, ..., 5.7010e-08, 3.6770e-07,
2.2495e-06]], device='cuda:0')}}

C:\Users\parth\Desktop\results>

```

Figure 1. DT-JSCC Accuracy [2, 8, 10] Results Using Classical Shannon Entropy Formulation.

```

Administrator: Command Prompt

03/31/2025  04:56 PM  <DIR>          .
03/31/2025  04:49 PM  <DIR>          ..
03/31/2025  04:54 PM                17,820 CIFAR10-num_e16-num_latent4-modpsk-psnr4.0-lam0.001.pt
03/30/2025  10:48 AM                161 rpytfile.py
                2 File(s)            17,981 bytes
                2 Dir(s)      789,619,187,712 bytes free

C:\Users\parth\Downloads\SB\SB\DT-JSSC\Discrete-TaskOriented-JSSC\results>.\rpytfile.py

C:\Users\parth\Downloads\SB\SB\DT-JSSC\Discrete-TaskOriented-JSSC\results>python rpytfile.py
{'acc1s': [87.890625, 88.75, 89.462890625, 89.287109375, 89.19921875, 89.70703125, 89.43359375, 89.23828125, 89.16015625,
89.345703125], 'acc3s': [97.294921875, 97.48046875, 97.8125, 97.8515625, 97.98828125, 97.98828125, 98.017578125, 97.85
15625, 98.037109375, 98.046875], 'dist': tensor([[9.9984e-01, 8.6570e-09, 2.8552e-11, ..., 6.1168e-09, 6.5997e-08,
1.9929e-07],
[1.0000e+00, 2.8362e-16, 5.8042e-20, ..., 2.3193e-22, 5.5863e-19,
1.6793e-16],
[1.0000e+00, 4.1113e-16, 1.6307e-19, ..., 1.1220e-23, 5.7181e-20,
4.0827e-17],
...,
[3.2576e-31, 3.4935e-23, 6.9299e-21, ..., 3.2136e-29, 3.8519e-29,
1.2040e-27],
[1.3787e-29, 5.2293e-20, 7.4686e-18, ..., 6.3587e-23, 1.1235e-23,
3.8687e-23],
[9.6542e-20, 5.8605e-13, 8.7258e-12, ..., 6.5195e-14, 8.7099e-15,
1.0443e-14]], device='cuda:0')}}

C:\Users\parth\Downloads\SB\SB\DT-JSSC\Discrete-TaskOriented-JSSC\results>

```

Figure 2. DT-JSCC Accuracy Results Using this Author's Variant of Shannon Entropy Formulation.

Experiment II

```

C:\Users\Partha Srinivasan\icer_compression\build>icer_util compress boatcolor.512.bmp output.bin --color
Loaded image: boatcolor.512.bmp (512x512, 3 channels)
Compression mode: Color (YUV)
Starting compression...
Parameters: stages=4, filter=0, segments=6, mode=lossless, quota=768.00KB
Compression completed in 0.278 seconds
Compressed size: 359240 bytes (45.7% of original)
Compressed image saved to: output.bin (359240 bytes)

C:\Users\Partha Srinivasan\icer_compression\build>dir/p
Volume in drive C is System
Volume Serial Number is 6264-A159

Directory of C:\Users\Partha Srinivasan\icer_compression\build

11/12/2025 05:25 AM <DIR> .
11/11/2025 04:27 PM <DIR> ..
11/11/2025 08:36 AM 786,486 boatcolor.512.bmp
11/11/2025 04:28 PM 14,926 CMakeCache.txt
11/11/2025 04:28 PM <DIR> CMakeFiles
11/11/2025 04:28 PM 1,815 cmake_install.cmake
11/11/2025 04:28 PM 311,439 compress.exe
11/12/2025 05:07 AM 99,977 compressed.bin
11/11/2025 04:28 PM 321,276 compress_color.exe
11/11/2025 04:28 PM 310,752 decompress.exe
11/11/2025 04:28 PM 320,589 decompress_color.exe
11/11/2025 04:28 PM 335,461 icer_util.exe
11/11/2025 04:28 PM <DIR> lib_icer
11/11/2025 04:28 PM 16,006 Makefile

```

Figure 3. IER Corresponding to Original Entropy Coder with "Probability of Zero" Principle.

```

Point 1: (0, 1, 32769)
Point 2: (1, 32769, 32770)
Point 1: (0, 0, 0)Point 2: (0, 0, 32768)SUM1073741824
Gets to 16
Gets here
2
D2
Created 2 embedded points, each with dimension 3.
Embedded data points:
Point 1: (1, 3, 32779)
Point 2: (3, 32779, 8)
Point 1: (0, 0, 0)Point 2: (0, 0, 0)SUM0
Compression completed in 29.810 seconds
Compressed size: 352452 bytes (44.8% of original)
Compressed image saved to: outg.bin (352452 bytes)

C:\Users\Partha Srinivasan\icer_compression\build>.\icer_util decompress outg.bin d.bmp --color
Error: Could not open input file outg.bin

C:\Users\Partha Srinivasan\icer_compression\build>.\icer_util decompress outg.bin d.bmp --color
Loaded compressed data: 352452 bytes
Image dimensions: 512x512
Starting decompression as color image...
Using parameters: stages=4, filter=0, segments=6
Decompression completed in 0.241 seconds
Decompressed to 512x512 color image
Decompressed image saved to: d.bmp

C:\Users\Partha Srinivasan\icer_compression\build>cd ..\build_ol

```

Figure 4. IER with Modified Entropy Criteria.

4. Conclusion

The entropy computation using traditional Shannon for-

mulation and that computed using the formulation introduced in this work yield almost identical accuracy in the evaluation of the recognition/classification task w.r.t. the data that is input at the source end and that which is received at receiver end.

The IER experiment provides yet another avenue to apply the notion of entropy to validate/substantiate existing results.

5. Future Works

It is envisaged that the alternative formulation of Entropy found in this work can potentially be used in several contexts and works where this statistic is employed resulting in aggregate refinement in understanding and application of this statistic.

Abbreviations

DT-JSCC Discrete Task Oriented Joint Source Channel Coding

Acknowledgements

The Author gratefully acknowledges the contributions made by the authors of the works cited herein and also the kind gesture of the publication for considering this effort.

Author Contributions

Parthasarathy Srinivasan is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Mutual Information-Empowered Task-Oriented Communication: Principles, Applications and Challenges Hongru Li, Songjie Xie, Jiawei Shao, Zixin Wang, Hengtao He Shenghui Song, Jun Zhang, and Khaled B. Letaief, <https://doi.org/10.48550/arXiv.2503.20195>
- [2] Robust Information Bottleneck for Task-Oriented Communication with Digital Modulation Songjie Xie, Student Member, IEEE, Shuai Ma, Member, IEEE, Ming Ding, Senior Member, IEEE, Yuanming Shi, Senior Member, IEEE, MingJian Tang, and Youlong Wu, Member, IEEE, <https://doi.org/10.48550/arXiv.2209.10382>
- [3] An Alternate Formulation of The Mutual Information Statistic Which Yields a More Realistic Measure, Leading to a More Precise and Dependable Model of Natural Language Translation Probabilities Among Parallel Corpora Author: Parthasarathy Srinivasan Beehive Software Solutions.
- [4] IPN Progress Report 42-155 November 15, 2003 The ICER Progressive Wavelet Image Compressor A. Kiely and M. Klimesh.
- [5] W. B. Pennebaker and J. L. Mitchell, JPEG Still Image Data Compression Standard, Van Nostrand Reinhold, New York, 1993. <https://archive.org/details/jpegstillimageda0000penn/page/n9/mode/2up>
- [6] A. R. Calderbank, I. Daubechies, W. Sweldins, and B.-L. Yeo, "Wavelet Transforms that Map Integers to Integers," Applied and Computational Harmonic Analysis, vol. 5, pp. 332-369, July 1998.
- [7] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," IEEE J. Sel. Areas Commun. vol. 41, no. 1, pp. 5-41, 2022.
- [8] C. Cai, X. Yuan, and Y.-J. A. Zhang, "End-to-end learning for task-oriented semantic communications over mimo channels: An information-theoretic framework," IEEE J. Sel. Areas Commun., 2025.
- [9] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, no. 3, pp. 379-423, 1948.
- [10] F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Joint source-channel coding for video communications," Handbook of Image and Video Processing, pp. 1065-1082, 2005.
- [11] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," IEEE Transactions on Signal Processing, vol. 69, pp. 2663-2675, 2021.
- [12] K. Wei, J. Li, C. Ma, M. Ding, C. Chen, S. Jin, Z. Han, and H. V. Poor, "Low-latency federated learning over wireless channels with differential privacy," IEEE Journal on Selected Areas in Communications, vol. 40, no. 1, pp. 290-307, 2021.
- [13] K. Choi, K. Tatwawadi, A. Grover, T. Weissman, and S. Ermon, "Neural joint source-channel coding," in International Conference on Machine Learning. PMLR, 2019, pp. 1182-1192. <https://proceedings.mlr.press/v97/choi19a.html>
- [14] Takens, F. (1981). Detecting strange attractors in turbulence. In: Rand, D., Young, L.S. (eds) Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics, vol 898. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/BFb0091924>