

Effat University Repository

Average entropy and tail heaviness of bell-shaped distributions: graphs and calculations

Item Type	article
Authors	Kittaneh, Omar
Citation	Kittaneh, O. (2025). Average entropy and tail heaviness of bell-shaped distributions: graphs and calculations. Communications in Statistics - Simulation and Computation, 1–13. https://doi.org/10.1080/03610918.2025.2565600
DOI	https://doi.org/10.1080/03610918.2025.2565600
Publisher	Informa UK Limited
Rights	Attribution-NonCommercial-NoDerivatives 4.0 International
Download date	2026-04-22 05:32:09
Item License	http://creativecommons.org/licenses/by-nc-nd/4.0/
Link to Item	https://repository.effatuniversity.edu.sa/handle/20.500.14131/2307

Average Entropy and Tail Heaviness of Bell-shaped Distributions: Graphs and Calculations

Omar Kittaneh*

*Professor and Director of STEM and CETL, STEM, College of Engineering, Effat University, 22332, Jeddah, Saudi Arabia

Email: omar_kittaneh@yahoo.com and okitanneh@effatuniversity.edu.sa

Abstract

This paper revisits the concept of average entropy, a functional measure of uncertainty introduced to address limitations in traditional entropy measures, particularly for continuous distributions. We investigate its utility as a consistent and interpretable measure of tail heaviness in bell-shaped probability distributions. Unlike kurtosis, which is undefined for many heavy-tailed distributions, average entropy is well-defined, location- and scale-invariant, and can be empirically estimated from data. We demonstrate that higher average entropy is associated with heavier distribution tails. This property is then used to define three canonical regions of the distribution: the peak, the shoulder, and the tail. The boundaries of these regions are identified in a consistent and unified manner across different distributions. These regions allow for objective comparisons of tail weight. The approach is applied to several well-known symmetric distributions, including the normal, student's t, exponential power, logistic, and Cramér distributions. Additionally, we demonstrate a practical estimation method using histogram-based approximations. The findings establish average entropy as a robust measure complementing kurtosis for characterizing tail behavior in symmetric distributions.

Keywords: Average entropy; Kurtosis; Bell-shaped distributions; Tail-heaviness.

1. Introduction

Measuring the tail heaviness of probability distributions is crucial in fields such as risk management, reliability engineering, and extreme value analysis, where the behavior of rare or extreme events can have significant consequences (Nair, Wierman, & Zwart, 2022).

However, many commonly used tail heaviness measures, such as kurtosis (Westfall, 2014), have well-documented limitations. Most importantly, kurtosis is undefined for many commonly used probability distributions, particularly those with heavy tails. Even when it is defined, kurtosis can be overly sensitive to outliers and may fail to offer consistent comparisons across different distribution families, as noted in earlier studies such as (Song, 2001) and more recent analyses like (Bastianin, 2019), (Eberl and Klar, 2025)

A consistent and reliable measure of tail heaviness, applicable across a wide range of probability distributions, is therefore essential. This work explores further properties of average entropy (Kittaneh et al., 2016) and extends its applications, proposing it as an intrinsic distributional measure that reflects tail heaviness in bell-shaped probability distributions from a perspective distinct from the traditional metric of kurtosis.

The average entropy, introduced by Kittaneh et al. (2016) and Kittaneh (2017), is a novel measure of uncertainty derived based on the concept of the average value. It was designed to address the limitations of Shannon entropy (1948), ensuring the same useful virtues for both discrete and continuous distributions. While Shannon entropy is guaranteed to be non-negative and invariant under linear transformation only for discrete distributions, these features do not hold for continuous ones. In contrast, continuous average entropy is a limit of discrete average entropy. As a result, average entropy can be reliably estimated from empirical distributions, but this is not possible for Shannon entropy.

The average entropy has been proposed as a potential clustering measure in image segmentation (Kittaneh, 2016; Kittaneh, 2023), although its application in image processing remains relatively underdeveloped.

The concept and usefulness of average entropy stem from the average value of the probability

density (or mass) function of the random variable in question. For a continuous random variable X with probability density function (PDF) f , the average value of f is defined as:

$$a_f = E_f[f(X)] = \int_a^b f^2(x)dx, \quad (1.1)$$

where (a, b) is its support.

The average entropy of a PDF f with support (a, b) is defined as

$$A_f = -E_f[\log\{f(X)/a_f\}] = -\int_a^b f(x)\log\{f(x)/a_f\}dx, \quad (1.2)$$

where “log” denotes the natural logarithm (i.e., logarithm to the base e), to avoid ambiguity. It can be shown that A_f has other equivalent forms to (1.2). The average entropy is the difference between

the Shannon entropy $H_f = -\int_a^b f(x)\log f(x)dx$, and the Renyi entropy (1961) of degree 2, $R_f = -\log \int_a^b f^2(x)dx$, i.e.,

$$A_f = H_f - R_f. \quad (1.3)$$

Equivalently, the average entropy is the difference between the logarithm of the expected value of f and the expected value of the logarithm of f , i.e.,

$$A_f = \log E_f[f(X)] - E_f[\log f(X)]. \quad (1.4)$$

It is also equal to the Kullback–Leibler divergence (1959) between f and its escort PDF (Girardin and Regnault, 2016) of degree 2 given by $g_f = f^2/a_f$. Specifically,

$$A_f = KL(f||g_f) = \int_a^b f(x)\log\{f(x)/g_f(x)\}dx. \quad (1.5)$$

The average entropy is always positive and equals zero when X is uniformly distributed. It is also a location and scale-invariant functional, meaning that

$$f(x) = \sigma^{-1}h((x - \mu)/\sigma), \text{ implies } A_f = A_h. \quad (1.6)$$

Section 2 reviews the concept of kurtosis, introduces its formula and discusses some of its drawbacks. Section 3 explores the relationship between tail heaviness and average entropy as an alternative to traditional kurtosis. Section 4 applies the entropy framework to common bell-shaped distributions and compares their tail behaviors. Section 5 shows how to estimate average entropy using histograms, with accuracy affected by sample size and tail weight. The paper is concluded in Section 6.

2. Kurtosis

Kurtosis has long been misinterpreted as a measure of a distribution's peakedness, flatness, pointiness, or modality. Its only clear interpretation is in relation to tail extremity and its ability to generate outliers (Westfall, 2014). For a random variable X with mean μ and standard deviation σ , the kurtosis of X is defined as:

$$K(X) = E[(X - \mu)^4]/\sigma^4 - 3. \quad (2.1)$$

However, this traditional measure of kurtosis (2.1) is undefined for a large class of probability distributions, including many bell-shaped distributions of heavy tails such as the Student's t , Cauchy, and Cramér distributions, which are symmetric and unimodal. It is also undefined for non-bell-shaped distributions like the log-logistic, Lévy, and Pareto distributions.

For example, the t -distribution with shape parameter (degrees of freedom) $\nu \leq 4$ has an undefined kurtosis. When $\nu = 1$, the t -distribution coincides with the Cauchy distribution, where even the mean does not exist. Using the traditional kurtosis measure, one cannot compare the Cauchy distribution to, say, the Laplace distribution, since the kurtosis for the Cauchy distribution is undefined. Similarly, one would mistakenly conclude that the Laplace distribution and the t -distribution with $\nu = 6$ have the same tail heaviness because both have a kurtosis of 3. The Cramér distribution presents a similar issue: it has a heavier tail than the Cauchy distribution, and its kurtosis is undefined. These examples highlight the limitations of kurtosis as a measure of tail heaviness in probability distributions, and more broadly, the limitations of using moments to describe the shape of a distribution.

3. The Average Entropy as a Measure of Tail Heaviness

According to equation (1.2), the average entropy is simply the expected value of the logarithm of the ratio between the density f and its average value. To understand how the average entropy relates to the heaviness of tails, we first consider the following results. Let g be a decreasing PDF on $[0, \infty)$, and define $f(x) = \frac{1}{2}g(|x|)$, which is a symmetric and unimodal density on $(-\infty, \infty)$, representing a general form of bell-shaped densities. In this case we have:

$$a_f = \frac{1}{2}a_g, \quad (3.1)$$

which is straightforward to verify: $a_f = \int_{-\infty}^{\infty} f^2(x)dx = 2 \int_0^{\infty} \frac{1}{4} g^2(|x|)dx = \frac{1}{2} \int_0^{\infty} g^2(x)dx = \frac{1}{2} a_g$.

It can also be shown that the average entropies of f and g are equal:

$$A_f = A_g. \quad (3.2)$$

This holds because $A_f = - \int_{-\infty}^{\infty} f(x) \log \left\{ \frac{f(x)}{a_f} \right\} dx = -2 \int_0^{\infty} \frac{1}{2} g(|x|) \log \left\{ \frac{g(|x|)}{a_g} \right\} dx = A_g$.

Therefore, to understand the average entropy of a bell-shaped density f , it is sufficient to study one of its branches. Let us now define the right and left truncated average entropies at x as

$$A_f(x) = -2 \int_0^x f(t) \log \left\{ \frac{f(t)}{a_f} \right\} dt, \quad (3.3)$$

$$\bar{A}_f(x) = -2 \int_x^{\infty} f(t) \log \left\{ \frac{f(t)}{a_f} \right\} dt, \quad (3.4)$$

respectively, where they are trivially related to the average entropy through the relationship

$$A_f = A_f(x) + \bar{A}_f(x). \quad (3.5)$$

It is clear that $A_f(x) < 0$ on the set $P_f = \{x: f(x) > a_f\}$. If f is continuous, there exists an x_a such that $P_f = (-x_a, x_a)$ with $f(-x_a) = f(x_a) = a_f$. Also, note that $\lim_{x \rightarrow \infty} A_f(x) = A_f$, which is a positive quantity, and since $A_f(x_a) < 0$ then there exists $x_A > x_a$ such that $A_f(x_A) = 0$. Thus in view of (3.5), this means that the average entropy is encapsulated entirely within the region of the density function defined on (x_A, ∞) . Specifically,

$$A_f = - \int_{T_f} f(t) \log \left\{ \frac{f(t)}{a_f} \right\} dt = -2 \int_{x_A}^{\infty} f(t) \log \left\{ \frac{f(t)}{a_f} \right\} dt = \bar{A}_f(x_A). \quad (3.6)$$

This is guaranteed by the intermediate value theorem as $A_f(x)$ is continuous. Now define $S_f = (-x_A, -x_a) \cup (x_a, x_A)$ and $T_f = (-\infty, -x_A) \cup (x_A, \infty)$. The three regions P_f , S_f and T_f are unique and are called the peak, shoulders, and tails regions induced by the average entropy of the density f .

In fact, on the right (or left) tail (x_A, ∞) we have

$$f(x) \leq f(x_A) \leq a_f \quad (3.7)$$

Additionally, for all the bell-shaped distributions we will discuss in Section 4, we observed that

$$f(x) \leq f(x_A) \leq \frac{a_f}{e}. \quad (3.8)$$

This suggests that, in general, higher values of average entropy typically indicate a greater value for the density on its tail. Although it is relatively straightforward to verify the validity of Equation (3.8) numerically for the distributions discussed in Section 4, it is still possible to derive some analytical insights. Let us define

$$L = f \log \frac{f}{a_f}, \quad (3.9)$$

Then $\frac{\partial L}{\partial f} = -1 - f \log \left\{ \frac{f}{a_f} \right\}$ which is nonnegative iff $f \leq \frac{a_f}{e}$.

The right and left truncated average entropies defined in Equations (3.3) and (3.4) decompose the global average entropy into contributions from specific regions of the distribution. Specifically, $A_f(x)$ captures the entropy from the central part (peak and shoulders), while $\bar{A}_f(x)$ reflects the contribution from the tail region. This localized perspective allows us to quantify where uncertainty is concentrated and to identify natural boundaries (e.g., x_a) that separate the core from the tails. Such a decomposition enhances our understanding of the distribution's shape and tail behavior, offering insights that complement the global entropy and are not captured by traditional measures like kurtosis. For more intuition about the behavior of $A_f(x)$, we mention that it always starts at zero and decreases becoming negative, and continues until it attains its minimum value, which happens at x_a , then increases and returns to zero at x_A , after that, it continues to increase and converge to the average entropy A_f . On the other hand, $\bar{A}_f(x)$ is always positive and starts at A_f , then increases to its maximum which happens at x_a then decreases back to A_f , which happens at x_A , and finally converges to zero.

4. Examples of Average Entropy in Common Bell-Shaped Distributions

In this section, we apply the proposed framework of average entropy to a variety of examples of commonly used bell-shaped probability distributions. These examples demonstrate how the average entropy provides insights into the structures of different distributions highlighting their peak, shoulder, and tail weights and regions, and contrast it with the traditional kurtosis-based interpretations.

Example 1 (Normal Distribution):

Figure 1 illustrates the concept proposed in Section 3 using the normal distribution. For a standard normal distribution with the PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}, x \in (-\infty, \infty) \quad (4.1)$$

The average value and average entropy are $a_f^N = \frac{1}{2\sqrt{\pi}}$ and $A_f^N = \frac{1}{2}(1 - \log 2) \approx 0.153$, respectively, where the superscript N denotes the normal distribution. It can also be verified that the positive solution of $f(x) = a_f^N$ is $x_a \approx 0.833$ and the solution of $A_f^N(x) = 0$ is $x_A \approx 1.799$.

In this case the peak, shoulders, and tails of f are defined as: $P_f = (-0.833, 0.833)$, $S_f = (-1.799, -0.833) \cup (0.833, 1.799)$, and $T_f = (-\infty, -1.799) \cup (1.799, \infty)$. Additionally, it's helpful to examine how the truncated average entropy $A_f^N(x)$, defined in equation (3.3) evolves as x increases, as shown in Figure 2.

The curve for $A_f^N(x)$ starts at zero, decreases to its minimum at $x_a \approx 0.833$, then rises back to zero at $x_A \approx 1.799$. Beyond x_A , the curve crosses above the x -axis and asymptotically approaches the average entropy $A_f^N \approx 0.153$. This behavior mirrors the shape of the PDF in Figure 1.

Table 1 presents the PDFs for a group of commonly used bell-shaped distributions, while Table 2 lists the corresponding average values and average entropies for these distributions.

Each distribution is shown in its standard form, with location parameters set to zero and scale parameters set to one. This standardization does not affect the results because of (1.6).

Having established the baseline behavior of the standard normal distribution, we now explore distributions with heavier tails to assess how average entropy reflects changes in tail weight.

Example 2 (Student's t -Distribution):

Our second example is the student's t -distribution, denoted by $t(\nu)$, where ν is the shape parameter. The PDF for this distribution is shown in Table 1, and its average value and average entropy are listed in Table 2.

When $\nu = 1$, the t -distribution reduces to the Cauchy distribution, denoted by CA .

The average entropy of the t -distribution decreases as ν increases. For instance, the approximate values of the average entropy are 0.693, 0.391, 0.303, 0.261, 0.238, and 0.222 for $\nu = 1, 2, 3, 4, 5$ and 6, respectively. Using asymptotic formulas for the gamma and digamma functions (Nemes, 2010), we can demonstrate that the average entropy for t -distribution, shown in Table 2, converges to $\frac{1}{2}(1 - \log 2) \approx 0.153$, which corresponds to the average entropy of the normal distribution.

In fact, when ν is large, the approximations $\Gamma[b + \nu] \approx \Gamma[\nu]\nu^b$ and $\psi[\nu] \approx \log(\nu - 0.5)$ hold, further supporting this convergence.

The student's t -distribution highlights how average entropy decreases with increasing degrees of freedom, aligning with the transition from heavy- to light-tailed behavior. Next, we consider the exponential power distribution to examine a more flexible family with varying tail heaviness.

Example 3 (Exponential Power Distribution):

The third example is the exponential power distribution, denoted by $EP(\delta)$, which has a shape parameter δ .

This distribution represents a broad class of bell-shaped distributions. As δ approaches ∞ , its PDF converges to

$$f(x) = \begin{cases} 0.5, & -1 < x < 1 \\ 0, & \text{ow} \end{cases} \quad (4.2)$$

which is the uniform distribution on $(-1,1)$. In fact, as $\delta \rightarrow \infty$, the average value of f converges to $1/2$ and its average entropy approaches zero, matching the average value and entropy, respectively for the uniform PDF on $(-1,1)$. In this case the distribution has no shoulders or tails—only a peak region that covers 100% of the area under the PDF.

While the exponential power distribution converges to the uniform distribution, which has the lightest tail of all bell-shaped distributions, it also generates heavier-tailed distributions.

For example, when $\delta = 2$, it produces the normal distribution, and for $\delta = 1$, it yields the Laplace distribution (LP), which has an even heavier tail. The tail of the Laplace distribution is comparable in weight to that of the t -distribution with $\nu = 2$. The Laplace distribution has a heavier tail than the normal distribution but a lighter tail than the Cauchy distribution.

Thus, the exponential power distribution serves as a bridge between distributions with extremely light and extremely heavy tails. We now turn to the logistic distribution to evaluate another commonly used distribution with symmetric bell shape but different tail characteristics.

Example 4 (Logistic Distribution):

The fourth example is the logistic distribution, denoted by *LG*. The average value of its density is $1/6$, which is the same as that of Cramer distribution, however, its average entropy is much smaller than the Cramer's average entropy. In fact, its average entropy is quite similar to that of the *t*-distribution with $\nu = 7$, and the weights of their tails are nearly identical, even though their x_a 's and x_A 's differ significantly. To further stress the distinction in tail weight beyond kurtosis or average value, we conclude this section with the Cramer distribution, which possesses exceptionally heavy tails.

Example 5 (Cramer Distribution):

In this example we discuss the Cramer distribution (*CR*), known for its heavy tails. While its average value is the same as that of the logistic distribution, its average entropy is significantly higher. In fact, the average entropy of the Cramer distribution is greater than that of the *t*-distribution for $\nu \geq 1$. Specifically, its average entropy of 0.901 exceeds that of the Cauchy distribution, which is 0.693, and its tail weight is also heavier, as will be discussed.

To visually reinforce the comparisons across the distributions analyzed above, we provide plots of their PDFs and truncated average entropy functions. These figures illustrate the characteristic shapes and thresholds derived from our proposed method. Figures 3, 4 and 5 depict the curves of the PDF, $f(x)$ and truncated AVE, $A_f(x)$ for the bell-shaped probability distributions illustrated in Examples 2 to 5. The horizontal red lines on the curves of the PDFs are for $f(x_a)$ and the blue lines are for $f(x_A)$. The vertical blue lines on both curves mark $\pm x_a$ and the red lines mark $\pm x_A$. The pattern seen in $A_f^N(x)$ in Figure 2 for the normal distribution is typical for all bell-shaped distributions, though the scales differ.

Table 3 lists the excess kurtosis, average entropy, and the weights of the peak, shoulders, and tails for the probability distributions we've discussed.

Table 3 compiles the key quantitative indicators excess kurtosis, average entropy, and region weights that help compare tail behaviors across distributions. These comparisons highlight the distinct advantages of average entropy as a descriptive metric. From Table 3, we observe that, generally, as kurtosis and average entropy increase, the tail weight $w(T)$ also increases, while the peak weight decreases. However, there is no clear relationship between these measures and the weight of the shoulders.

Unlike kurtosis, which is undefined for many heavy-tailed distributions like the t -distribution with $\nu \leq 4$ and Cramer distribution, average entropy is defined for all probability distributions. Moreover, the average entropy not only measures tail heaviness but also defines the regions of the distribution's tail, peak, and shoulders as discussed in Section 3. Finally, to complete the characterization of distributional shapes, we provide the explicit intervals corresponding to peak, shoulder, and tail regions based on the entropy-based thresholds. Table 4 provides these regions for the distributions listed in Table 3.

Example 6 (Discrete Laplace Distribution):

Unlike the previous examples, which focus on continuous distributions, this final example examines a discrete case; specifically, the Discrete Laplace Distribution introduced by Inusah and Kozubowski (2006). This distribution has the probability mass function (PMF)

$$p(x) = \frac{1-p}{1+p} p^{|x|}, \text{ where } x \in Z = \{\dots, -2, -1, 0, 1, 2, \dots\} \quad (4.3)$$

where $p = e^{-1}$. The primary motivation behind the original works by Kittaneh et al. (2016) and Kittaneh (2017) was to develop an entropy measure that applies consistently to both discrete and continuous distributions, preserving parallel properties in both domains.

For a probability mass function $p(x)$, the discrete average entropy is defined in exactly similar way as the continuous average entropy as

For a given PMF $p(x)$ with support $\{x_1, x_2, x_3, \dots\}$ without loss of generality, the discrete average entropy is defined analogously to the continuous case as

$$A_p = -E_p[\log\{p(x)/a_p\}] = -\sum_{x=x_1}^{\infty} p(x) \log \frac{p(x)}{a_p} \quad (4.4)$$

where a_p is the average value of the PMF $p(x)$, given by

$$a_p = \sum_{x=x_1}^{\infty} p^2(x) \quad (4.5)$$

Moreover, the continuous average entropy can be viewed as the limit of the discrete version, as demonstrated in Theorem 4 of Kittaneh et al. (2016). Consequently, the behavior of the discrete case closely mirrors that of the continuous one, with the main distinction being that the cutoff points x_a and x_A might not lie within the discrete support.

It can be numerically verified that $a_p \approx 0.28$ and $A_p \approx 0.35$, which are remarkably close to the values for the continuous Laplace distribution $a_f \approx 0.25$ and $A_f \approx 0.307$. For the discrete Laplace case, the approximate cutoff points are $x_a \approx 0.5$ and $x_A \approx 1.1$. These results are illustrated in Figure 6.

5. Histogram Estimation of the Average Entropy

As we've seen, the average entropy can either be constant or depend on unknown parameters that typically need to be estimated. A more practical and interesting question arises: given a set of independent observations from a density f , without knowing the exact form of the distribution, how can we estimate the average entropy?

Since the average entropy cannot be directly expressed as a function of the distribution function F , nonparametric estimation using the empirical distribution function is not feasible. However, Kittaneh et al. (2016) proved in Theorem 4 that average entropy can be estimated using histograms of empirical data, regardless of the parent distribution. It can also be estimated through kernel density estimators, where Pinsker (1964) demonstrated that the Kullback-Leibler measure of kernel densities converges to the true value in general, and in particular for average entropy, as shown in representation (1.5). In this work, we will focus on the simpler method of histogram estimation.

Table 5 presents the histogram-based estimates of average entropy using the Freedman and Diaconis (1981) binning method, based on Monte Carlo samples of size $n = 10, 50, 100, 1000$, generated from the distributions discussed earlier. This binning scheme was selected because it outperforms other methods, such as Sturges (Scott, 2009) or Scott (Scott, 2010) binning methods.

These estimates were obtained by following the steps (i–viii) outlined below:

- i. Generate a random sample of size n from the distribution of interest f .

- ii. Determine the bins b_i using the method of Freedman and Diaconis (1981).
- iii. Compute the weight of each bin, p_i , by counting the number of specimens inside each bin.
- iv. Calculate the Shannon entropy $\hat{H}_f = -\sum_i p_i \log p_i$.
- v. Calculate the Renyi entropy of degree 2, $\hat{R}_f = -\log \sum_i p_i^2$.
- vi. Determine the average entropy, $\hat{A}_f = \hat{H}_f - \hat{R}_f$.
- vii. Calculate the mean absolute percentage error (MAPE) $e_f = |A_f - \hat{A}_f|/A_f$ (de Myttenaere, Golden, Le Grand, & Rossi, 2016), where A_f is the actual average entropy of the distribution f . Any other relative error would work. Note that A_f is always nonnegative and is equal to zero if and only if f is a uniform density which has no tails. This note has been added.
- viii. Repeat steps (i-vii) $M = 5000$ times and compute the mean values of quantities (vi) and (vii).

From Table 5, it's evident that the expected value of the estimated average entropy converges to the true value. Generally, the smallest errors occur for the normal distribution or distributions with average entropies close to that of the normal distribution. This is because binning methods are designed based on the normal distribution. The errors increase significantly when the average entropy is either very large (indicating a heavy tail) or very small (indicating a light tail). In such cases, a larger sample size is needed to achieve the desired level of accuracy.

6. Comments and Future Directions

This work proposes the use of average entropy as a measure of tail heaviness, though it does not suggest replacing the traditional measure, kurtosis. While kurtosis may be undefined for many probability distributions, it has the advantage of being easily estimated from random samples, with statistics that converge to the true value when kurtosis is defined. In contrast, average entropy is applicable to nearly all distributions, but its estimation, as suggested in this work, requires preprocessing steps such as constructing histograms or using kernel methods. These processes can affect the quality of estimation and generally require large sample sizes for reliable results. Although the Freedman and Diaconis binning method outperforms Sturges and Scott methods, all

provide similar estimates of average entropy for the normal distribution and for distributions with average entropies that are not significantly higher or lower than that of the normal distribution.

This study also introduces a method for defining the peak, shoulder, and tail regions for bell-shaped distributions based on average entropy. While some may disagree with the specific intervals, the method is applied uniformly across all bell-shaped distributions, allowing for a consistent and fair ordering (\preceq) of distributions based on tail heaviness. The proposed order is:

$$\text{Exp. Power}(\delta \geq 2) \preceq \text{Normal} \preceq t(v \geq 3) \preceq \text{Laplace} \preceq t(v = 2) \preceq \text{Cauchy} \preceq \text{Cramer},$$

which we believe most would probably agree to this order.

The illustration in this work is only for continuous distributions; however, the findings might be valid for discrete distributions with some modifications.

Acknowledgments

All results, figures, tables and discussions are solely developed by the main author.

Disclosure Statement

The authors report there are no competing interests to declare.

References

- [1] Bastianin, A. (2019). Robust measures of skewness and kurtosis for macroeconomic and financial time series. *Applied Economics*, 52(7), 637–670.
- [2] de Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing*, 192, 38–48.
- [3] Eberl, A., Klar, B. (2025). Measures of kurtosis: inadmissible for asymmetric distributions?. *Metrika* **88**, 365–391.
- [4] Freedman, D., & Diaconis, P. (1981). On the histogram as a density estimator: L2 theory. *Probability Theory and Related Fields*, 57(4), 453–476.
- [5] Girardin, V., & Regnault, P. (2016). Escort distributions minimizing the Kullback–Leibler divergence for a large deviation principle and tests of entropy level. *Annals of the Institute of Statistical Mathematics*, 68, 439–468.

- [6] Kittaneh, O., Khan, M. A., Akbar, M., & Bayoud, H. (2016). Average entropy: A new uncertainty measure with application to image segmentation. *American Statistician*, 70, 18-24.
- [7] Kittaneh, O. (2017). Response to average entropy does not measure uncertainty. *American Statistician*, 71(1), 1.
- [8] Kittaneh, O., "The Generalized Average Entropy with Applications to some Satellite Image Thresholding," 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 2023, pp. 1-4
- [9] Inusah, S., & Kozubowski, T. J. (2006). A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136(3), 1090–1102.
- [10] Kullback, S. (1959). *Information theory and statistics*. Wiley.
- [11] Nair, J., Wierman, A., & Zwart, B. (2022). *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge University Press. Online ISBN: 9781009053730.
- [12] Nemes, G. (2010). New asymptotic expansion for the gamma function. *Arch. Math. (Basel)*, 95, 161–169.
- [13] Pinsker, M. S. (1964). *Information and information stability of random variables and processes*. Holden-Day. (Originally published in Russian in 1960).
- [14] Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 547-561.
- [15] Scott, D. W. (2009). Sturges' rule. *WIREs Computational Statistics*, 1(3), 303–306.
- [16] Scott, D. W. (2010). Scott's rule. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4).
- [17] Shannon, C. E. (1948). A mathematical theory of communications. *Bell System Technical Journal*, 27, 379-423.

- [18] Song, K.-S. (2001). Rényi information, loglikelihood and an intrinsic distribution measure. *Journal of Statistical Planning and Inference*, 93(1–2), 51–69.
- [19] Westfall, P. H. (2014). Kurtosis as peakedness, 1905–2014. R.I.P. *The American Statistician*, 68(3), 191–195.

Figures

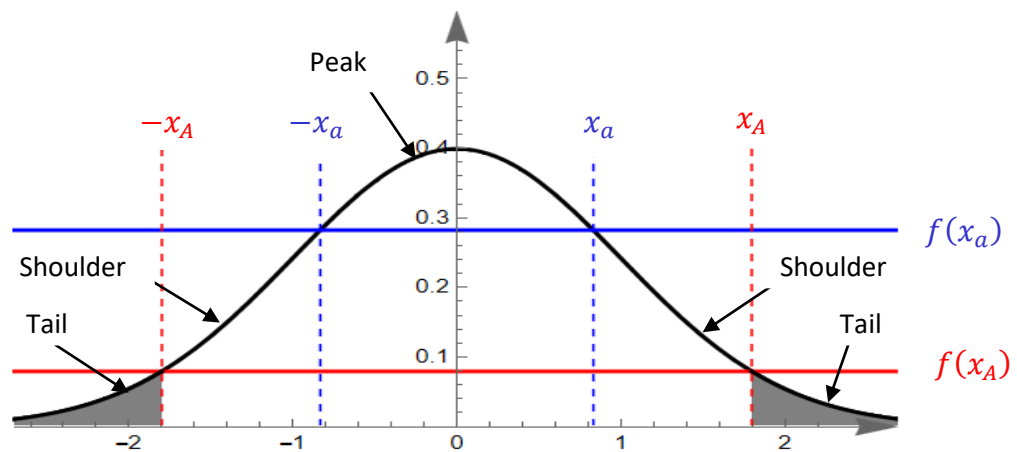


Figure 1. The standard normal PDF with its peak, shoulders, and tails.

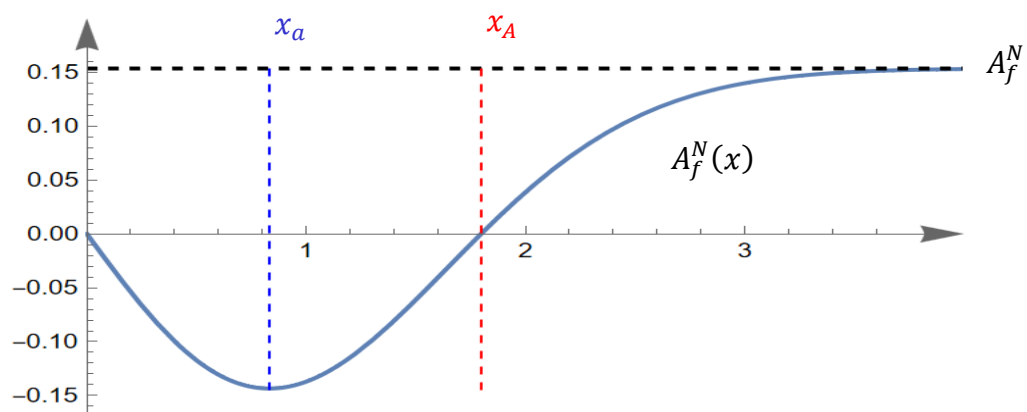


Figure 2. The truncated average entropy $A_f^N(x)$ curve and the average entropy A_f^N for the standard normal distribution.

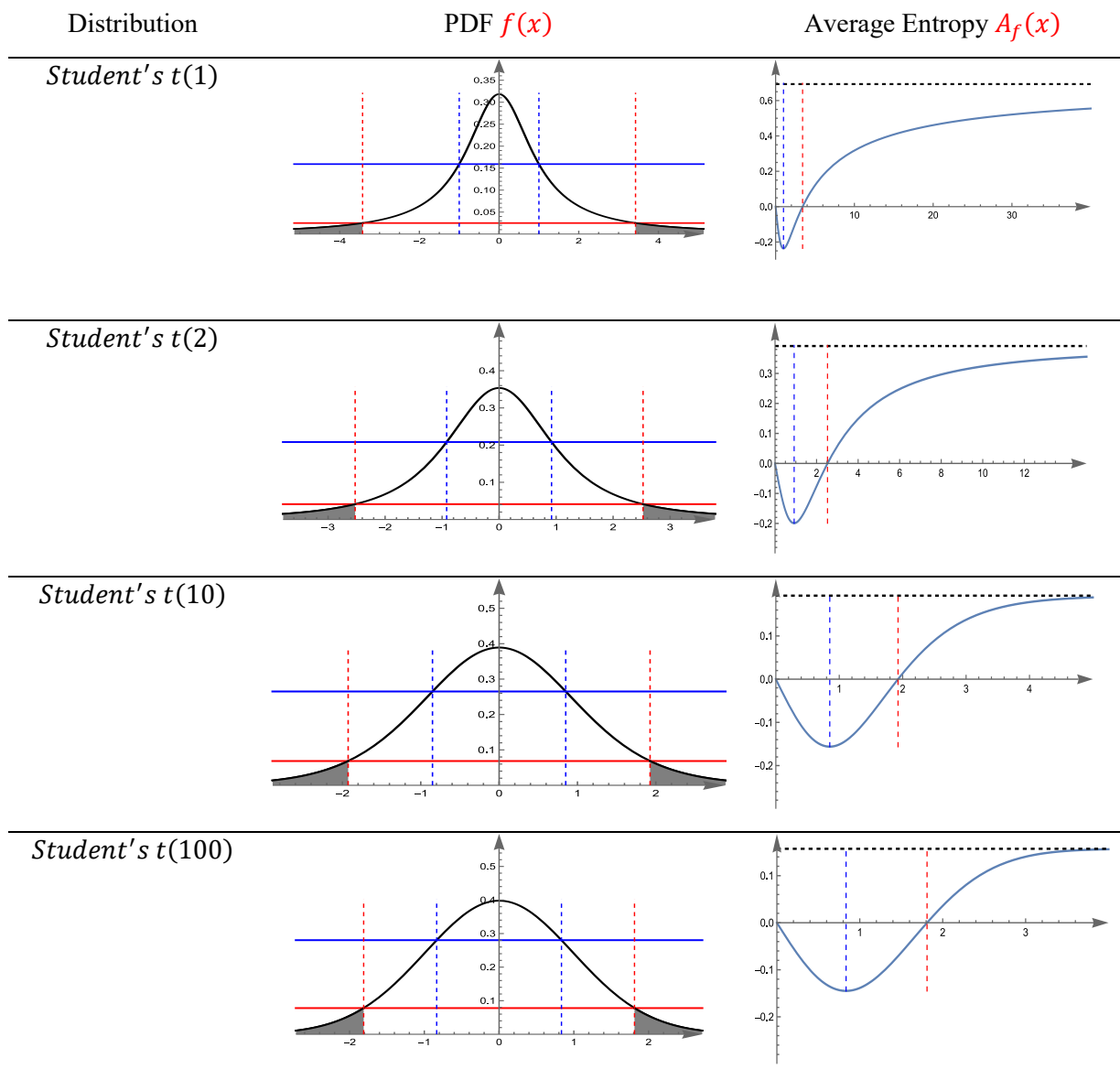


Figure 3. The PDF and AVE for several student's t distributions

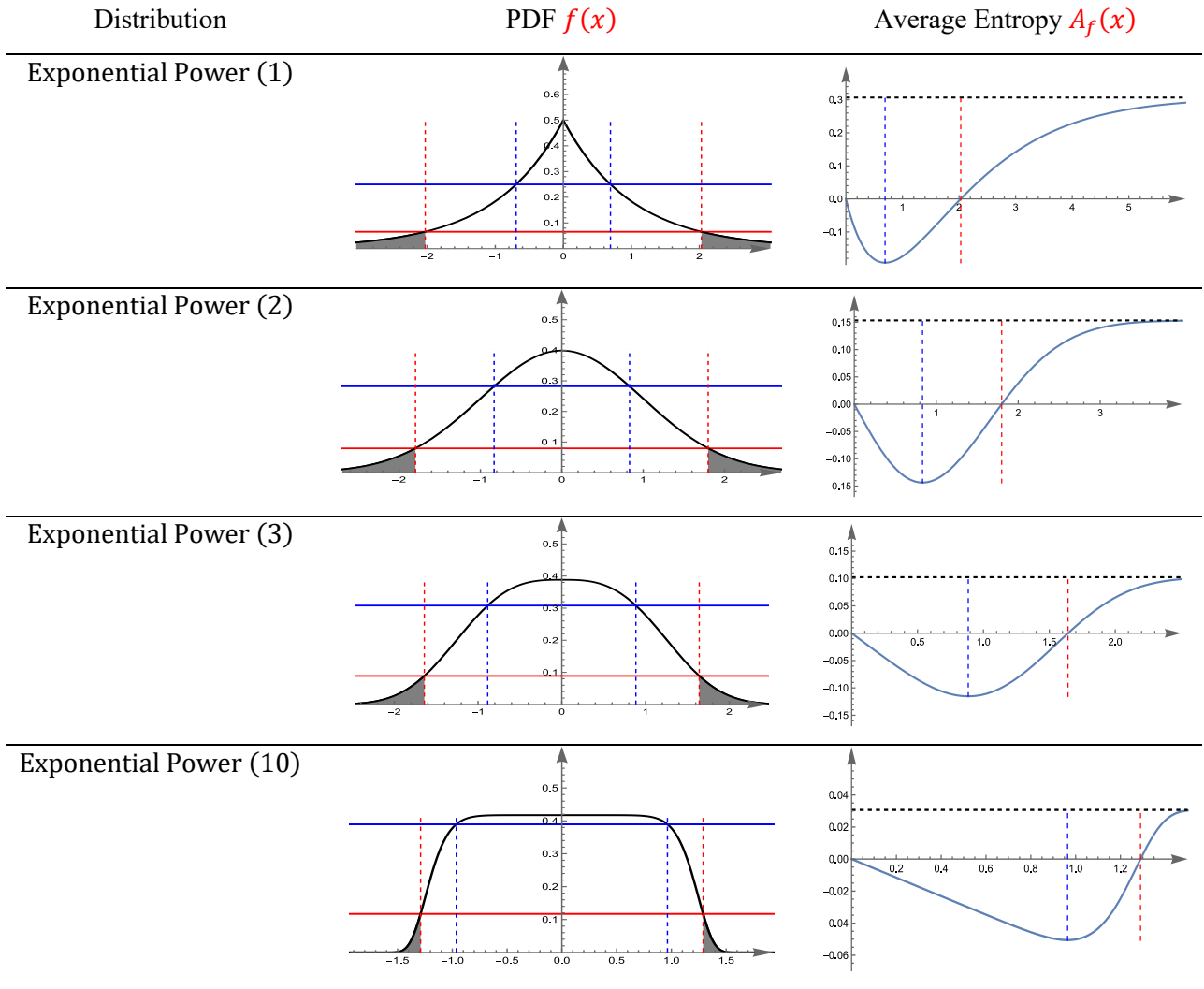


Figure 4. The PDF and AVE for several exponential power distributions

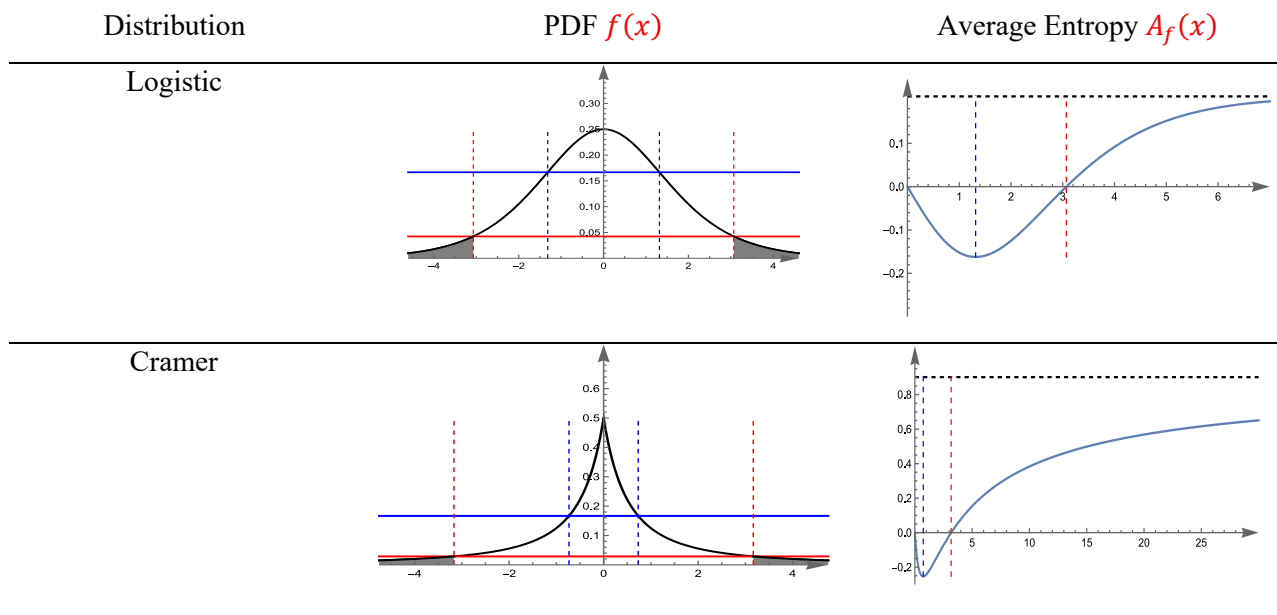


Figure 5. The PDF and AVE for logistic and Cramer distributions

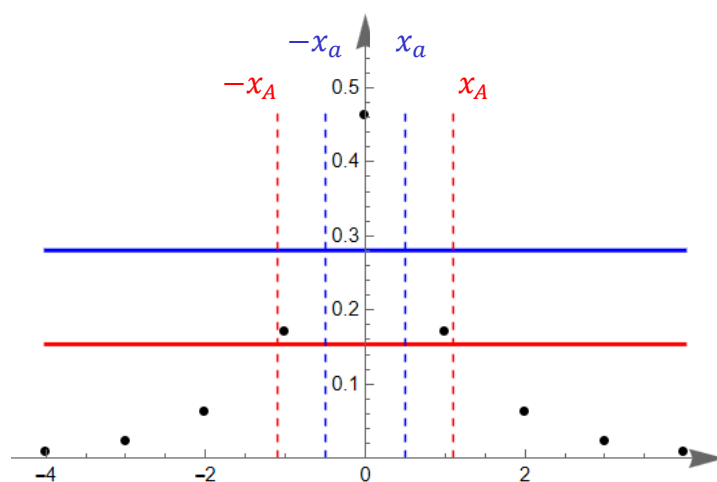


Figure 6. The PMF for discrete Laplace distribution

Figure captions (as a list).

Figure 1. The standard normal PDF with its peak, shoulders, and tails

Figure 2. The truncated average entropy $A_f^N(x)$ curve and the average entropy A_f^N for the standard normal distribution

Figure 3. The PDF and AVE for several student's t distributions

Figure 4. The PDF and AVE for several exponential power distributions

Figure 5. The PDF and AVE for logistic and Cramer distributions

Figure 6. The PMF for discrete Laplace distribution

Tables

Table 1: Frequently used bell-shaped distributions on $(-\infty, \infty)$

Distribution	PDF f
Exponential Power (δ)	$\frac{\delta^{-1/\delta}}{2\Gamma[1+1/\delta]} \exp\left\{\frac{-1}{\delta} x ^\delta\right\}, \delta > 0$
<i>Normal</i>	$\frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-1}{2}x^2\right\}$
<i>Student's t</i> (ν)	$* \frac{\Gamma[0.5(\nu+1)]}{\sqrt{\pi\nu}\Gamma[0.5\nu]} \left(1 + \frac{x^2}{\nu}\right)^{-0.5(\nu+1)}, \nu > 0$
<i>Logistic</i>	$\frac{\exp\{-x\}}{(1 + \exp\{-x\})^2}$
<i>Laplace</i>	$\frac{1}{2} \exp\{- x \}$
<i>Cauchy</i>	$\frac{1}{\pi(1 + x^2)}$
<i>Cramer</i>	$\frac{1}{2(1 + x)^2}$

* Γ is the gamma function

Table 2: Average value and average entropy for the distributions in Table 1

Distribution	a_f	A_f
Exponential Power (δ)	$\frac{1}{2(2\delta)^{1/\delta}\Gamma[1+1/\delta]}$	$\frac{1}{\delta}(1-\log 2)$
<i>Normal</i>	$\frac{1}{2\sqrt{\pi}}$	$\frac{1}{2}(1-\log 2)$
<i>Student's t</i> (ν)	$\frac{\Gamma[0.5+\nu]\Gamma[0.5(\nu+1)]^2}{\sqrt{\pi}\nu^{1.5}\Gamma[0.5\nu]^2\Gamma[\nu]}$	$\log\left\{\frac{\mathcal{B}[1/2+\nu,1/2]}{\mathcal{B}[\nu/2,1/2]}\right\}^*$ $+ \frac{(1+\nu)}{2}\left(\psi\left[\frac{\nu+1}{2}\right]-\psi\left[\frac{\nu}{2}\right]\right)^{**}$
<i>Logistic</i>	1/6	$2-\log 6$
<i>Laplace</i>	1/4	$1-\log 2$
<i>Cauchy</i>	$1/(2\pi)$	*** $-\gamma-\log 2-\psi[1/2]\approx 0.693$
<i>Cramer</i>	1/6	$2-\log 3$

* \mathcal{B} is the beta function, ** ψ is the digamma function, *** γ is the Euler's constant ≈ 0.577

Table 3: Kurtosis vs Average Entropy

Distribution	Kurtosis	Average Entropy	* $w(P)$	* $w(S)$	* $w(T)$
$EP(\delta = 15)$	-1.159	0.020	0.84	0.15	0.01
$EP(\delta = 10)$	-1.116	0.031	0.80	0.18	0.02
$EP(\delta = 5)$	-0.930	0.061	0.72	0.25	0.03
$EP(\delta = 4)$	-0.812	0.077	0.69	0.27	0.04
$EP(\delta = 3)$	-0.582	0.102	0.65	0.30	0.05
N	0.000	0.153	0.60	0.33	0.07
$t(\nu = 100)$	0.063	0.157	0.60	0.33	0.07
$t(\nu = 10)$	1.000	0.193	0.59	0.33	0.08
LG	1.200	0.208	0.58	0.33	0.09
$t(\nu = 7)$	2.000	0.212	0.58	0.33	0.09
$t(\nu = 6)$	3.000	0.222	0.58	0.33	0.09
$t(\nu = 5)$	6.000	0.238	0.57	0.33	0.10
$t(\nu = 4)$	**UD	0.261	0.57	0.33	0.10
$t(\nu = 3)$	UD	0.303	0.56	0.33	0.11
LP	3.000	0.307	0.50	0.37	0.13
$t(\nu = 2)$	UD	0.391	0.54	0.33	0.13
CA	UD	0.693	0.50	0.32	0.18
CR	UD	0.901	0.42	0.34	0.24

* $w(P)$, $w(S)$ and $w(T)$ denote the weights of the peak, shoulders, and tails, respectively. **UD:

undefined

Table 4: Peak: $-x_a < x < x_a$, shoulders: $-x_A < x < -x_a$ and $x_a < x < x_A$, and tails: $x < -x_A$ and $x > x_A$

Distribution	x_a	x_A
$EP(\delta = 15)$	0.976	1.215
$EP(\delta = 10)$	0.964	1.290
$EP(\delta = 5)$	0.929	1.467
$EP(\delta = 4)$	0.912	1.539
$EP(\delta = 3)$	0.885	1.642
N	0.833	1.799
$t(\nu = 100)$	0.834	1.811
$t(\nu = 10)$	0.850	1.927
LG	1.317	3.069
$t(\nu = 7)$	0.858	1.985
$t(\nu = 6)$	0.862	2.018
$t(\nu = 5)$	0.868	2.064
$t(\nu = 4)$	0.877	2.136
$t(\nu = 3)$	0.891	2.261
LP	0.834	1.811
$t(\nu = 2)$	0.920	2.526
CA	1.000	3.425
CR	0.732	3.166

Table 5: Histogram estimation for the average entropy, accurate to 3 decimal places, based on Freedman and Diaconis binning method

Distribution	A_f	$\hat{A}_f(e_f)$			
		$n = 10$	$n = 50$	$n = 100$	$n = 1000$
$EP(\delta = 15)$	0.020	0.074(2.855)	0.050(1.702)	0.064(2.148)	0.034(0.682)
$EP(\delta = 10)$	0.031	0.076(1.822)	0.056(1.633)	0.066(1.185)	0.040(0.297)
$EP(\delta = 5)$	0.061	0.084(0.793)	0.080(0.560)	0.078(0.357)	0.066(0.106)
$EP(\delta = 4)$	0.077	0.087(0.607)	0.090(0.395)	0.089(0.281)	0.080(0.083)
$EP(\delta = 3)$	0.102	0.094(0.491)	0.107(0.296)	0.107(0.212)	0.104(0.068)
N	0.153	0.103(0.428)	0.140(0.246)	0.145(0.179)	0.151(0.059)
$t(\nu = 100)$	0.157	0.104(0.435)	0.141(0.244)	0.147(0.180)	0.154(0.058)
$t(\nu = 10)$	0.193	0.107(0.480)	0.159(0.249)	0.170(0.183)	0.186(0.061)
LG	0.208	0.110(0.496)	0.168(0.252)	0.181(0.185)	0.200(0.059)
$t(\nu = 7)$	0.212	0.109(0.498)	0.170(0.259)	0.181(0.189)	0.202(0.064)
$t(\nu = 6)$	0.222	0.112(0.506)	0.174(0.263)	0.186(0.198)	0.211(0.066)
$t(\nu = 5)$	0.238	0.113(0.530)	0.182(0.272)	0.197(0.204)	0.224(0.070)
$t(\nu = 4)$	0.261	0.115(0.562)	0.192(0.288)	0.209(0.216)	0.243(0.076)
$t(\nu = 3)$	0.303	0.120(0.601)	0.210(0.314)	0.232(0.247)	0.276(0.089)
LP	0.307	0.121(0.608)	0.219(0.3000)	0.243(0.219)	0.289(0.066)
$t(\nu = 2)$	0.391	0.126(0.677)	0.245(0.372)	0.277(0.294)	0.344(0.121)
CA	0.693	0.143(0.791)	0.336(0.513)	0.403(0.419)	0.549(0.208)
CR	0.901	0.149(0.824)	0.385(0.575)	0.479(0.469)	0.703(0.221)