ABSTRACT

A COMPARISON OF VARIANCE AND RENYI'S ENTROPY WITH APPLICATION TO MACHINE LEARNING

Adric Peccarelli, MS
Division of Statistics
Northern Illinois University, 2017
Dr. Nader Ebrahimi, Director

This research explores parametric and nonparametric similarities and disagreements between variance and the information theoretic measure of entropy, specifically Renyi's entropy. A history and known relationships of the two different uncertainty measures is examined. Then, twenty discrete and continuous parametric families are tabulated with their respective variance and Renyi entropy functions ordered to understand the behavior of these two measures of uncertainty. Finally, an algorithm for variable selection using Renyi's Quadratic Entropy and its kernel estimation is explored and compared to other popular selection methods using real data.

NORTHERN ILLINOIS UNIVERSITY
DEKALB, ILLINOIS

AUGUST 2017

A COMPARISON OF VARIANCE AND RENYI'S ENTROPY WITH APPLICATION TO MACHINE LEARNING

BY

ADRIC M. PECCARELLI

A THESIS SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

DIVISION OF STATISTICS

Thesis Director:
Dr. Nader Ebrahimi

ACKNOWLEDGEMENTS

I would like to express my appreciation for many individuals. Firstly, my thesis director, Dr. Nader Ebrahimi for his continued enthusiasm into my area of focus and this research, and allowing me to take this research in directions I deemed personally interesting. Secondly, Dr. Duchwan Ryu for helping me develop a method to quickly analyze my parametric formulations. Thirdly, I would like to thank the professors and my fellow graduate students in the Northern Illinois University Division of Statistics for their support and motivational inspiration. Lastly, I would like to thank my family, friends, and Brittany Rios, for their continued tolerance and support in regards to my education.

TABLE OF CONTENTS

Pages

LIST OF TABLES………………………………………………………………………….. iv

LIST OF FIGURES………………………………………………………………………… v

LIST OF APPENDICES…………………………………………………………………… vi

Chapter

1. Variance and Renyi's Spectra of Information…………………………………… 1

Introduction………………………………………………………………… 1

Variance and Entropy………………………………………………………. 1

Renyi's Spectra of Information…………………………………………….. 2

Discussion………………………………………………………………….. 4

2. Parametric Comparison…………………………………………………….. 5

Distribution Ordering by Variance and Entropy……………………………. 5

Notation…………………………………………………………………….. 5

Location-scale Families…………………………………………………….. 5

Shape-scale Families………………………………………………………… 7

Student t, F, and Beta Families……………………………………………. 10

Discrete Distributions………………………………………………………… 12

Discussion…………………………………………………………………. 14

3. Nonparametric Comparison of Variance and Entropy……………………………. 15

Chapter Page

Information Theoretic Learning…………………………………………….. 15

ITL Framework………………………………………………………… 15

Variable Selection Algorithm……………………………………………….. 17

Dataset……………………………………………………………………. 19

Entropy Estimation for Variable Selection………………………………….. 19

Methods of Comparison……………………………………………………. 20

Results……………………………………………………………………. 21

Discussion………………………………………………………………… 23

4. Conclusion……………………………………………………………………… 25

REFERENCES………………………………………………………………………… 26

APPENDICES…………………………………………………………………………...27

LIST OF TABLES

Table Page

1. Variance and Renyi's Entropy Ordering for Location-scale Families……………….. 6

2. Variance and Renyi's Entropy Ordering for Shape-scale Families………………….. 8

3. Variance and Renyi's Entropy Ordering For Student t, F, and Beta Distributions…... 11

4. Variance and Renyi Entropy Ordering for Discrete Distributions…………………... 13

5. Method Statistics……………………………………………………………………….. 21

LIST OF FIGURES

Figure Page

1. Density of Target Prediction Accuracy…………………………………………………… 22

2. Comparison Between the True Target Response and Method Estimations…………. 23

LIST OF APPENDICES

Appendix Page

A. RENYI’S QUADRATIC ENTROPY KERNEL DENSITY ESTIMATION R CODE… 27

B. ORDERING EXAMPLE R CODE……………………………………………………….. 29

1. VARIANCE AND RENYI SPECTRA OF INFORMATION

1.1 Introduction

In many scientific disciplines, variance as a measure of uncertainty is prevalent. The general computational simplicity contributes to the popularity of this measure in roles such as evaluating dispersion of data and goodness of fit, and the assessment and ranking of random variables. While there is no question that variance is successful and plays an important part in these roles, many works in recent decades as discussed in Ebrahimi et al. (1999, 2010) suggest there are other superior measures of information, particularly the entropy classes found in Information Theory.

The area of Information Theory was first popularized by Claude Shannon with “A Mathematical Theory of Communication” (1948). Since that publication many areas of study, including statistics, have adapted uses for information theoretic measures, with one such measure being entropy, or the measure of uncertainty in some state.

1.2 Variance and Entropy

There are a few intuitive similarities and differences between variance and entropy to discuss. They are both measures of concentration or dispersion under some state. An important distinction between the two is that variance is a measure centered only around the density mean, while entropy is a measure which is irrespective of its locations within the density. Another difference is that entropy measures the divergence of a density from the uniform distribution, which holds maximum uncertainty, while variance is an average of distances from the mean of its own distribution. This can be rationalized as the entropy of a distribution always having a consistent baseline irregardless of the distribution, where variance is dependent of its own distribution.

Aside from the intuitive relationship between variance and entropy, there are many mathematical properties that can be used to describe both. For a discrete random variable, both variance and entropy are non-negative, but only entropy is invariant under one-to-one transformations. For a continuous random variable, entropy can take values across the real line while variance is still non-negative. Neither are invariant under one-to-one transformations in this case.

For continuous random variables, $\mathrm{E}(\mathrm{X}^2) < \infty \implies \mathcal{H}(X) < \infty$, but the converse may not hold. Shannon proposed a comparison between the entropy and variance of a continuous random variable,

$$\frac{e^{2\mathcal{H}(X)}}{2\pi e} \leq Var(X), \tag{1}$$

where equality holds if and only if the random variable X is normal. This is known as the entropy power fraction and shows an early emphasis into the relationship between variance and entropy.

While there are notable differences between entropy and variance, the overall similarity in their measure of uncertainty within a density is enough to warrant further investigation into their relationship.

1.3 Renyi Spectra of Information

There are many entropy measures that have been developed since Shannon's original publication, with one such measure being the Spectrum of Renyi Information of Order λ, or Renyi's entropy.

Let X be a continuous (discrete) random variable on $\mathbb{R}^n$ with probability density (mass) function p(x), then differential Renyi entropy is defined as:

$$\mathcal{H}_{(\mathrm{R},\lambda)}(x) = \tfrac{1}{1-\lambda} log(\textstyle\int_{\mathbb{S}} p^{\lambda}(x)dx) \tag{2}$$

and the Renyi entropy for discrete random variables is defined as:

$$\mathcal{H}_{(\mathrm{R},\lambda)}(x) = \frac{1}{1-\lambda} log(\Sigma_{k=1}^{n} p^{\lambda}(x_k)) \tag{3}$$

where $\lambda > 0, \lambda \neq 1$. The value λ can be interpreted as a leveraging point, which can give different weights to different probabilities. As $\lambda \to 0$, all events, regardless of their respective probabilities, are weighed more equally. As $\lambda \to \infty$, the entropy becomes more dependent on the events with highest probabilities.

The well known result for $\lambda = 1$ is,

$$\mathcal{H}_{(\mathrm{R},1)}(x) = \lim_{\lambda \to 1} \mathcal{H}_{(\mathrm{R},\lambda)}(x) = -\int_{\mathbb{S}} p(x) log(p(x)) dx \tag{4}$$

which is the widely used differential Shannon entropy. The same result follows for the discrete case. Therefore, Renyi's entropy is the known generalization of Shannon entropy. Having the logarithm on the outside of the integral as with Renyi's measure as opposed to Shannon's leads to some unique results that will be discussed further on.

While there are many properties of entropy, Shannon stated four important requirements, or axioms:

1) *Continuity:* $\mathcal{H}(x)$ is continuous in $x \in \mathbb{R}^n$
2) *Symmetry*: $\mathcal{H}(x) = \mathcal{H}(x_1, x_2, ... x_n)$ is invariant under permutations of $x, i = 1, ...n$
3) *Monotonicity*: $\mathcal{H}(x)$ is a monotonically increasing function
4) *Partition Invariance:* For a pair of random variables with joint pdf $\mathrm{p}(\mathrm{x}_1, x_2)$,

$$\mathcal{H}(p(x_1, x_2)) = \mathcal{H}(p(x_1)) + \int_{\mathbb{S}} \mathcal{H}(p(x_2|x_1)) dP(x_1) \tag{5}$$

Alfred Renyi (1961) proposed an alternative condition to axiom 4, which is the weaker additivity under independence

$$\mathcal{H}(p(x_1, x_2)) = \mathcal{H}(p(x_1)) + \mathcal{H}(p(x_2)) \tag{6}$$

Axiom 4 clearly results in this condition under independence and this alternative leads to the Spectrum of Renyi Information of Order λ. The remainder of this research will focus on this measure of entropy.

1.4 Discussion

This research will expand on Ebrahimi et al. (1999), which looked at the ordering of many popular continuous and discrete distributions by variance and Shannon entropy to try to establish patterns between the two measures. In Section 2, we will look at the natural extension of that paper by repeating their process using the generalization of Shannon's measure, Renyi's entropy, to explore where the previously established results agree or if any disagreements arise. Following that in Section 3, we will develop a Renyi entropy based variable selection and discuss its implications to machine learning, as well as compare this concept to other selection methods. Overall conclusions will follow in Section 4.

2. PARAMETRIC COMPARISON OF VARIANCE AND ENTROPY

2.1 Distribution Ordering by Variance and Entropy

In this section, we will look at sixteen continuous parametric families and four discrete distributions and tabulate their variance and Renyi's entropy, along with a directional relationship between the two measures with respect to λ and their familial parameters $(\alpha, \beta) \in \theta$. The continuous parametric families are separated into the same groups as presented in Ebrahimi et al.. These four family groups are: *Location-scale families; Shape-scale families; Student t, F, and Beta families; and Discrete Distributions*.

2.2 Notation

The notation used in this section will follow from Ebrahimi et al.(1999). Let Ω represent the class of probability distributions for which the variance and entropy are under consideration. For two random variables X_1 and X_2 with distributions $\mathrm{P}_1 \in \Omega$ and $\mathrm{P}_2 \in \Omega$, respectively, the variances and entropies of the distributions will be denoted as $\mathrm{V}_1, \mathrm{V}_2$ and $\mathrm{H}_1, \mathrm{H}_2$.

Let Ω_θ represent a class of distributions indexed by parameter $\theta \in \mathbb{R}_\theta$. We will denote when variance, in which $\mathrm{V}(\mathrm{X}|\theta) < \infty$, is increasing (decreasing) in θ as $\mathrm{V} \nearrow \theta$ ($\mathrm{V} \searrow \theta$). We will denote when entropy, in which $|\mathrm{H}(\mathrm{X}|\theta)| < \infty$, is increasing (decreasing) in θ as $\mathrm{H} \nearrow \theta$ ($\mathrm{H} \searrow \theta$). Also for entropy, we will look at the effect the constant λ has on the ordering. Since $\lambda \neq 1$, we will be interested when $\lambda \in (0, 1)$ and when $\lambda > 1$. When the two measures order similarly, we will use $(\mathrm{V}, \mathrm{H}) \nearrow \theta$ or $(\mathrm{V}, \mathrm{H}) \searrow \theta$.

2.3 Location-scale Families

For any distribution with density in the form of:

$$p(x|\alpha, \beta) = \frac{1}{\beta}\mathrm{p}(\frac{x - \alpha}{\beta})$$

Table 1: Variance and Renyi's Entropy Ordering for Location-scale Families

Family	Density $p(x)$	Variance	Renyi Spectra of Information $\mathcal{H}_{R,\lambda}(x)$	Ordering
Gaussian (Normal)	$\frac{1}{\sqrt{2\pi}\beta}e^{[-\frac{1}{2\beta^2}(x-\alpha)^2]}$	β^2	$\frac{1}{2}log(2\pi\beta^2) - \frac{1}{2}(1-\lambda)^{-1}log(\lambda)$	$(V,H) \nearrow \beta$
Gumbel (Extreme Value)	$\frac{1}{\beta}e^{[-\frac{(x-\alpha)}{\beta} - exp(-\frac{(x-\alpha)}{\beta})]}$	$\frac{\pi^2\beta^2}{6}$	$log(\beta) - \frac{\lambda}{1-\lambda}log(\lambda) +$ $(1-\lambda)^{-1}log(\Gamma(\lambda))$	$(V,H) \nearrow \beta$
Laplace (Double Exponential)	$\frac{1}{2\beta}exp[-\frac{\lvert x-\alpha\rvert}{\beta}]$	$2\beta^2$	$\frac{1}{2}log(2\beta^2) - (1-\lambda)^{-1}log(\lambda)$	$(V,H) \nearrow \beta$
Logistic	$\frac{1}{\beta}\frac{exp(-\frac{(x-\alpha)}{\beta})}{(1+exp(\frac{(x-\alpha)}{\beta}))^2}$	$\frac{\pi^2\beta^2}{3}$	$\frac{1}{2}log(\beta) - (1-\lambda)^{-1}log\mathrm{B}(\lambda,\lambda)$	$(V,H) \nearrow \beta$
Uniform	$\frac{1}{\beta}$ $\alpha - \frac{\beta}{2} < x < \alpha + \frac{\beta}{2}$	$\frac{\beta^2}{12}$	$log(\beta)$	$(V,H) \nearrow \beta$

with location parameter α and scale parameter β, this distribution is said to be in the location-scale family. For the results presented in Table 1, all of the densities have been parameterized such that the location parameter is α and the scale parameter is β.

Well-known location-scale families of distributions were selected: Gaussian and continuous Uniform for their overall importance and common use; Gumbel for its use in analyzing maximum sample values; Laplace for its use in Bayesian regression and machine learning; and Logistic for its usefulness for modeling categorical data.

As mentioned in Ebrahimi et. al.(1999), variance and entropy for all location-scale distributions are independent of the location parameter α, which can be seen in the results from Table 1. A continuous random variable X from a location-scale distribution can be written as $X = \beta Z + \alpha$. Therefore, the variance and entropy for random variable X is:

$$\mathrm{V}(X|\alpha,\beta) = \beta^2\mathrm{V}(Z)$$

and

$$\mathcal{H}(X|\alpha,\beta) = \mathrm{Log}\beta + \mathcal{H}(Z)$$

where $\mathrm{V}(Z)$ and $\mathcal{H}(Z)$ are constants independent of α and β. From this and the results presented show that $(\mathrm{V},\mathrm{H}) \nearrow \beta$ for a fixed λ. For a fixed β, when $\lambda \in (0,\infty)$ and $\lambda \neq 1$, $\mathcal{H}_{(\mathrm{R},\lambda)}(\mathrm{X})$ decreases as λ increase $\forall\lambda$.

For location-scale families of distributions, we can see that some distinct patterns emerge. Both variance and entropy are independent from the location parameter α and increase in β for any fixed λ. The last observation for a fixed β is interesting for the entropy. Given a fixed value of β, in order to minimize the loss of information for location-scale distributions, all that is necessary is to increasingly adjust the weight parameter λ.

2.4 Shape-scale Families

For any distribution with density of the form:

$$p(x|\alpha,\beta) = \frac{1}{\beta}\mathrm{p}(\frac{x}{\beta}|\alpha)$$

with shape parameter α and scale parameter β, this distribution is said to be in the shape-scale family. For the results presented in Table 2, all of the densities have been parameterized such that the shape parameter is α and the scale parameter is β.

Some well-known and widely used shape-scale families of distributions were selected. The Gamma distribution, along with its special cases the Exponential and Chi-Squared distributions, are widely used amongst multiple disciplines. The Log-normal distribution is widely used for describing natural growth and is therefore applicable in many fields dealing with human biology and behavior. The Pareto distribution is applicable for situations dealing with extreme differences, such as economics and income dispersion, internet traffic, and natural

Table 2: Variance and Renyi's Entropy Ordering For Scale-shape Families

Family	Density $p(x)$	Variance	Renyi Spectra of Information $\mathcal{H}_{R,\lambda}(x)$	Ordering
Gamma; Exponential, α=1; Erlang, α=1,2,...; Chi-square, $\alpha=\frac{1}{2},\frac{2}{2},...\beta=2$	$\frac{X^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}e^{(-\frac{x}{\beta})}$	$\beta^2\alpha$	$log(\beta) + (1-\lambda)^{-1}log(\Gamma^{\lambda}(\alpha)) - log(\lambda)\frac{\lambda\alpha+(1-\lambda)}{1-\lambda}$	$(V,H)\nearrow\beta$ $H\nearrow\alpha$ $\forall\lambda\in(0,1)$, $H\searrow\alpha$ $\forall\lambda>1$
Inverse Gamma; Inverse Chi-square, $\alpha=\frac{1}{2},\frac{2}{2},...\beta=\frac{1}{2}$	$\frac{1}{\beta\Gamma(\alpha)}(\frac{\beta}{x})^{\alpha+1}e^{(-\frac{\beta}{x})}$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \quad \alpha>2$	$log(\beta)+(1-\lambda)^{-1}log[\frac{\Gamma(\lambda\alpha+(\lambda-1))}{\lambda^{(\lambda+\lambda\alpha-1)}\Gamma^{\lambda}(\alpha)}]$	$(V,H)\nearrow\beta$ $(V,H)\searrow\alpha>2$ $\forall\lambda$
Generalized-Normal; Half-Normal, α=1; Rayleigh, α=2; Maxwell-Boltzmann, α=3; Chi, α=1,2,...	$\frac{1}{2\beta\Gamma(\frac{\alpha}{2})}(\frac{x}{\beta})^{\alpha-1}e^{(-\frac{x}{\beta})^2}$	$\frac{\beta^2[\alpha-2\Gamma^2(\frac{\alpha}{2}+\frac{1}{2})]}{2\Gamma^2(\frac{\alpha}{2})}$	$\frac{1}{1-\lambda}[-\lambda log(2\beta\Gamma(\frac{\alpha}{2}))+(\frac{-\alpha\lambda+\lambda-1}{2})log(-\frac{1}{2}\beta\lambda)+log(\Gamma(\frac{1}{2}((\alpha+1)\lambda+1),\frac{x^2\lambda}{\beta^2})^{**}]$	V $\nearrow\alpha,\beta$ $H\nearrow\alpha\searrow\beta$ $\lambda\in(0,1)$, $H\searrow\alpha\nearrow\beta$ $\lambda>1$
Inverse Generalized Normal; InverseChi, $\alpha=1,2,...\beta=\sqrt{\beta}$	$\frac{1}{2\beta\Gamma(\frac{\alpha}{2})}(\frac{\beta}{x})^{\alpha+1}e^{(-\frac{\beta}{x})^2}$	$\beta^2[\frac{1}{\alpha-2}-\frac{\alpha\Gamma^2(\alpha-1)}{\Gamma^2(\alpha/2)}]$	$\frac{1}{1-\lambda}[-\lambda log(2\beta\Gamma(\frac{\alpha}{2}))+(\frac{1-\alpha\lambda+\lambda}{2})log(\frac{1}{2}\beta\lambda)+log(\Gamma(\frac{1}{2}((\alpha+1)\lambda-1),\frac{\beta^2\lambda}{x^2})^{**}]$	V $\nearrow\beta\searrow\alpha$ $H\searrow\alpha,\beta$ $\lambda\in(0,1)$, $H\nearrow\alpha,\beta$ $\lambda>1$ $\alpha>2$
Log-Normal	$\frac{1}{\sqrt{2\pi}\alpha x}e^{(-\frac{1}{2\alpha^2}(log(\frac{x}{\beta}))^2}$	$\beta^2[e^{2\alpha^2}-e^{-\alpha^2}]$	$\frac{(1+\lambda)^2\alpha^2}{2\lambda(1-\lambda)}-\frac{\alpha^2}{1-\lambda}-\frac{log(\lambda)}{2(1-\lambda)}+\frac{log(\frac{2\pi\alpha^2}{\beta^2})}{2}$	V $\nearrow\alpha,\beta$ $H\nearrow\alpha\searrow\beta$ $\lambda\in(0,1)$, $H\searrow\alpha\nearrow\beta$ $\lambda>1$
Pareto	$\frac{\alpha}{\beta}(\frac{x}{\beta})^{-(\alpha+1)}$, $x>\beta$	$\frac{\beta^2}{\alpha(\alpha-2)^2(\alpha-2)}, \quad \alpha>2$	$\frac{\lambda log(\alpha)}{1-\lambda}-\frac{log[(1+\alpha)\lambda-1]}{1-\lambda}+log(\beta)$	$(V,H)\searrow\alpha$ $(V,H)\nearrow\beta$ $\forall\lambda \quad \alpha>2$
Triangular	$\{\begin{matrix}\frac{2}{\alpha\beta}(\frac{x}{\beta}), & 0\leq x\leq\alpha\beta \\ \frac{2}{(1-\alpha)\beta}(1-\frac{x}{\beta}), & \alpha\beta\leq x\leq\beta\end{matrix}$	$\frac{\beta^2(\alpha^2-\alpha+1)}{18}$	$log(\beta)+\frac{log(\frac{2\lambda}{\lambda+1})}{1-\lambda}$	V $\nearrow\alpha,\beta$ $H\nearrow\beta$ (H constant $\forall\alpha$) $\forall\lambda$
Weibull	$\alpha(\frac{x}{\beta})^{\alpha-1}exp(-(\frac{x}{\beta})^{\alpha})$	$\beta^2[\Gamma(1+\frac{2}{\alpha})-\Gamma^2(1+\frac{1}{\alpha})]$	$\frac{log(\Gamma(\lambda-\frac{\lambda-1}{\alpha})}{1-\lambda}-\frac{\lambda log(\lambda)}{\lambda-1}-log(\frac{\alpha}{\beta}\lambda^{\alpha})$	V $\searrow\alpha\nearrow\beta$ $H\nearrow\alpha\searrow\beta$ $\forall\lambda$

*where ** is the Incomplete Gamma function:* $\Gamma(\alpha^*, \beta^*) = \int_{\beta^*}^{\infty} t^{\alpha^*-1} e^{-t} dt$

disaster assessment. The Weibull distribution has application in areas where describing a "time to failure" is important, such as survival analysis and reliability control.

Along with these important scale-shape families, some lesser used distributional families were also selected. The Triangular distribution applies to a population in situations where limited data is available, using a maximum and minimum. The Generalized-normal is of interest when the special case of both high value concentration around the mean and tail behavior are being considered. The Inverse Gamma has applications in Bayesian analysis, and finally the Inverse Generalized-normal has use in describing tendencies in Brownian motion.

From the results in Table 2, it is apparent that the Renyi's entropy for these Scale-shape distributions are complicated functions with little distinct patterns between them. The results in Ebrahimi et al.(1999) still hold for most of these families for a portion of the spectrum of λ, but generally not for all λ. For example, Gamma was found to both increase in α and β for both variance and Shannon entropy. For Renyi's entropy though, H $\nearrow \alpha \; \forall\lambda \in (0, 1)$, but decreases $\forall\lambda > 1$.

When λ is analyzed around the limit though, the parameters for all presented families begin to coincide with the previous results. The Generalized-normal and Log-normal distributions demonstrate identical behavior of both the variance and entropy across all λ, which is understandable given their descriptive relation to each other. An interesting behavior of the Inverse Generalized-normal when $\lambda > 1$ is that the directions of the variance and entropy reverse given the parameters. The Inverse Gamma, Pareto, Triangular, and Weibull distributions do not depend on the value of λ and behave similarly to previous results for most values of the parameters, so defaulting to Shannon entropy for computational simplicity is acceptable.

While there are no simple generalizing formulations in order to write the variance and Renyi entropies for these scale-shape families as there was with the presented location-scale families, having these measures expressed in a concise table will hopefully be beneficial to those in fields where these distributions are useful.

2.5 Student t, F, and Beta Families

This section will look at three well known and important continuous distributions, which would not directly fall into the previously discussed families. Table 3 parameterizes these distributions in the same previously used α, β notation.

The important observation to make for these distributions is that with Renyi's measure results in significantly less complicated functions when compared to Shannon's measure found in Ebrahimi et al.(1999). This is most noticable with the Beta distribution, whose Shannon entropy resulted in an ordering that required multiple planes in a three dimensional space to explain the behavior of the parameters. In our case here, the scaling factor λ in Renyi's measure greatly simplifies the expression and helps explain the overall direction the entropy exhibits despite the disruption that occurs at the limit $\lambda \to 1$.

The remaining distributions in Table 3 exhibit another interesting effects when observing the spectra of λ. For the student t distribution, the behavior of the variance and Shannon entropy $((V, H) \searrow \alpha)$ correspond to the results in Table 3 when $\lambda \in (0, 1)$, but entropy increases for $\lambda > 1$. The converse of this is true for the F distribution in regards to Shannon entropy. The behavior of the variance and Shannon entropy $(V \searrow \alpha, H \nearrow \alpha)$ corresponds for $\lambda > 1$, but

Table 3: Variance and Renyi's Entropy Ordering For Student t, F, and Beta Distributions

Family	Density $p(x)$	Variance	Renyi Spectra of Information $\mathcal{H}_{R,\lambda}(x)$	Ordering
Student-t; α=1,2,... Chaucy; α=1	$p(x\|\alpha) =$ $\frac{1}{\sqrt{\pi\alpha}}(\frac{\Gamma(\frac{\alpha}{2}+\frac{1}{2})}{\Gamma(\frac{\alpha}{2})})(1+\frac{x^2}{\alpha})^{-\frac{(\alpha+1)}{2}}$	$\frac{\alpha}{\alpha-2}$, $\alpha > 2$	$\frac{1}{1-\lambda}log[\frac{\mathrm{B}(\frac{\alpha\lambda}{2}+\frac{\lambda}{2}-\frac{1}{2},\frac{1}{2})}{\mathrm{B}^{\lambda}(\frac{\alpha}{2},\frac{1}{2})}] + log(\alpha)$	$V \searrow \alpha$ $H \searrow \alpha$ $\lambda \in (0,1)$, $H \nearrow \alpha$ $\lambda > 1$ $\alpha > 2$
F; α=1,2,...β=1,2,...	$p(x\|\alpha,\beta) =$ $\frac{\alpha^{\alpha/2}\beta^{\beta 2}\mathrm{x}^{\alpha/2-1}}{\mathrm{B}(\frac{\alpha}{2},\frac{\beta}{2})(\beta+\alpha\mathrm{x})^{(\alpha+\beta)/2}}$	$\frac{2\beta^2}{\alpha(\alpha+\beta-1)(\beta-2)(\beta-4)}$, $\beta > 4$	$\frac{1}{1-\lambda}log[\frac{\mathrm{B}(\frac{\lambda(\alpha-2)}{2}+1,\frac{\lambda(\beta+2)}{2}-1)}{\mathrm{B}^{\lambda}(\frac{\alpha}{2},\frac{\beta}{2})}]$ $-log(\frac{\alpha}{\beta})$	$V \searrow \alpha$ $H \searrow \alpha$ $\lambda \in (0,1)$ $H \nearrow \alpha$ $\lambda > 1$, $(V,H) \searrow \beta > 4$ $\forall\lambda$
Beta; $\alpha > 0$, $\beta > 0$	$p(x\|\alpha,\beta) =$ $\frac{\mathrm{x}^{\alpha-1}(1-\mathrm{x})^{\beta-1}}{\mathrm{B}(\alpha,\beta)}$, $0 \leq \mathrm{x} \leq 1$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$	$\frac{1}{1-\lambda}log[\frac{\mathrm{B}(\lambda(\alpha-1)+1,\lambda(\beta-1)+1)}{\mathrm{B}^{\lambda}(\alpha,\beta)}]$	$(V,H) \searrow (\alpha,\beta)$ $\forall\lambda$

entropy decreases when $\lambda \in (0,1)$. Both entropy measures behave the same with respect to the parameter β. When observing at the limit, there is an opposition in the behavior of the entropy, but it can be seen that the student t and F distributions demonstrate the same behavior for variance and Renyi's entropy in regards to the parameter α.

While with close examination of all the previous distributions in this research, the relationship of the behavior between the two measures can be observed, the distributions in Table 3 are important within the scope of this research to quickly visualize and interpret the ability of

Renyi's measure to detect the true nature of the entropy of a distribution and the "disruption" that occurs when $\lambda \to 1$.

2.6 Discrete Distributions

This section explores the relationship between the discrete Binomial, Geometric, Poisson, and Uniform distributions. While in the previous sections we were implementing the differential Renyi's entropy formula (2) to compute our functions, here we will use formula (3) to compute our functions for these discrete distributions. The typical lexicon for these distribution's parameters is used in Table 4.

For the Geometric and Poisson distributions we see similar entropic behavior between Shannon and Renyi's measures as was observed in the previous section. The variance and Renyi's entropy of the Geometric and Poisson distribution correspond with the results for Shannon's entropy described in Ebrahimi et al.(1999) for $\lambda \in (0, 1)$, but the entropy changes direction for $\lambda > 1$. In these examples, one could describe the relationship of Shannon and Renyi's entropy for the Geometric and Poisson distributions as:

$$\mathcal{H}_S(x) \simeq \mathcal{H}_R(x)$$

$$\text{when} \quad \lambda \in (0, 1].$$

The Binomial distribution is known to be symmetric around $p = \frac{1}{2}$, and this is apparent with the variance which is increasing (decreasing) for $V \nearrow p < \frac{1}{2}$ $(V \searrow p > \frac{1}{2})$. When examining the Shannon entropy for the Binomial, this symmetry is also apparent with the entropy and variance corresponding in ordering. The Renyi entropy does not depend on this symmetry, and only the

Table 4: Variance and Renyi Entropy Ordering for Discrete Distributions

Family	Density $p(x)$	Variance	Renyi Spectra of Information $\mathcal{H}_{R,\lambda}(x)$	Ordering
Binomial(n,p)	$p(x\|n,p) = \binom{n}{x} p^x(1-p)^{n-p}$ $x = 0, 1, \ldots n$	$np(1-p)$	$\frac{1}{1-\lambda} log(1-(1-p)^{n\lambda})$	$V \nearrow p < \frac{1}{2}$ $H \searrow \forall p$ $\lambda \in (0,1)$ $H \nearrow \forall p$ $\lambda > 1$
Geometric(p)	$p(x\|p) = p(1-p)^x$ $x = 0, 1, \ldots$	$\frac{1-p}{p^2}$	$\frac{\lambda}{1-\lambda} log[(p-1)((1-p)^n - 1)]$	$V \searrow p$ $H \searrow p$ $\lambda \in (0,1)$, $H \nearrow p$ $\lambda > 1$
Poisson(θ)	$p(x\|\theta) = \frac{e^{\theta}\theta^x}{x!}$ $x = 0, 1, \ldots$	θ	$\frac{\lambda}{1-\lambda}[log(\frac{\Gamma(n+1,\theta)}{\Gamma(n+1)}) - e^{-\theta}]$	$V \nearrow \theta$ $H \nearrow \theta$ $\lambda \in (0,1)$ $H \searrow \theta$ $\lambda > 1$
Uniform(n)	$p(x\|n) = \frac{1}{n}$, $x = 1, 2, \ldots n$	$\frac{n^2-1}{12}$	$log(n)$	$(V, H) \nearrow n$

position of λ. The Renyi entropy also provides a closed and more concise function as compared to Shannon's measure.

The ordering and Renyi entropy of the discrete Uniform distribution is equivalent to the continuous interpretation in Table 1 given the respective parameters and is also equivalent to the respective Shannon entropy functions. This is a natural conclusion since the uniform holds the maximum uncertainty of any distribution, and is "considered the *global reference distribution* for quantifying information in terms of predictability" (Ebrahimi, Soofi, & Soyer, 2010). Therefore, it should be consistent across all entropy measures in both discrete and continuous cases, regardless of any parameters such as the scaling factor in Renyi's measure.

2.7 Discussion

This chapter examined connections between the variance and entropy of twenty continuous and discrete families of distributions, focusing on Renyi's measure of entropy and, in some cases, comparing the directional behavior in regards to Shannon's measure of entropy.

The variance and entropy of Location-scale families of distributions can be expressed in distinct generalize functions. Patterns for Shape-scale families of distributions, as a whole, are much harder to express, and any distinctions can usually only be determined between two or more certain distributions. Tables 3 and 4 demonstrates how the spectra of Renyi's scaling factor λ helps to compare Renyi's measure to Shannon's. From this, we were able to simplify these functions of entropy and interpret the true entropic behavior for these distributions when $\lambda \to 1$.

This chapter also presents a comprehensive and concise collection of variance and Renyi entropy functions for the multiple families of distributions. Having a collection of these expressions for the presented distributions should be beneficial for researchers across many disciplines.

3. NONPARAMETRIC COMPARISON OF VARIANCE AND ENTROPY

3.1 Information Theoretic Learning

The areas of data mining and machine learning have become increasingly researched and implemented across most major scientific and social disciplines in recent years. While these large areas of research are well developed through typical statistical methods, a growing number of researchers, including Jose C. Principe and colleagues, are applying information theoretic methods in lieu of more accepted variance based measures.

"Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives" (Jose C. Principe, 2010) collects and presents many information theory methods while comparing them to the more traditional supervised and unsupervised learning algorithms.

The remainder of this research will set up the framework for Information Theoretic Learning using Renyi's entropy, apply this framework to a variable selection algorithm in terms of data compression, and then compare the performance to the Lasso and Backward/Forward variable selection.

3.2 ITL Framework

The ITL framework presented in [2] begins with *Renyi's Quadratic Entropy* ($\lambda = 2$):

$$\begin{aligned}\mathcal{H}_{(\mathrm{R},2)}(x) &= \tfrac{1}{1-2} log(\textstyle\int_{\mathbb{S}} p^2(x)dx) \\ &= -log(\textstyle\int_{\mathbb{S}} p(x)p(x)dx) \\ &= -logE[p(x)] \end{aligned} \tag{7}$$

The expectation of the PDF (PMF in the discrete case) is referred to as the information potential in the literature, and will be denoted as $V_2(x) = E[p(x)]$.

In most cases, one is left to estimate entropy directly from the data nonparametrically. Instead of estimating the PDF first and then calculating the entropy, when using Renyi's Quadratic Entropy, one just has to estimate the scalar $V_2(x)$.

Suppose for a continuous random variable X, we have N independent and identically distributed samples $\{x_1, x_2, ..., x_N\}$. The well-known estimate of the PDF with an arbitrary kernal $k_\sigma(.)$ is given by

$$\hat{p}_X(x) = \frac{1}{N\sigma}\Sigma_{i=1}^{N} k(\frac{x - x_i}{\sigma}) \tag{8}$$

where σ is the bandwidth parameter.

From here we can assume a Gaussian kernel to find our estimate for our quadratic entropy:

$$\hat{\mathcal{H}}_{(\mathrm{R},2)}(x) = -log \int_{-\infty}^{\infty} (\tfrac{1}{N}\Sigma_{i=1}^{N} G_\sigma(x - x_i))^2 dx$$

$$= -log\frac{1}{N^2}\int_{-\infty}^{\infty}(\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} G_\sigma(x - x_j)G_\sigma(x - x_i))dx$$

$$= -log\frac{1}{N^2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N}\int_{-\infty}^{\infty} G_\sigma(x - x_j)G_\sigma(x - x_i)dx$$

$$= -log[\frac{1}{N^2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{N} G_{\sigma\sqrt{2}}(x_j - x_i)]$$

$$= -log[\hat{V}_2(x)] \tag{9}$$

Principe (2010) states that this framework provides the most convenient expression for this entropy estimate, although other positive definite kernels that peak around the origin can still be used.

Now we are just estimating the information potential of Renyi's quadratic entropy directly using the data instead of the PDF. While this is similar to estimating the sample mean and variance to understand the shape of the distribution, the information potential still has the bandwidth parameter to estimate within the function. Principe (2010) suggests that a sufficient estimate for the bandwidth here follows "Silverman's Rule":

$$\sigma_{opt} = \sigma_X (4N^{-1}(2d+1)^{-1}))^{\frac{1}{(d+4)}} \tag{10}$$

where N is the number of samples, σ_X is the data standard deviation, and d is the dimensionality of the data.

3.3 Variable Selection Algorithm

The last concept outside the ITL framework to review before describing our variable selection algorithm is mutual information. Mutual information (MI) measures the mutual dependence of two variables, while entropy is a measure of uncertainty for a single random variable (C.T. Liu et al, 2010).

Given two random variables, X and Y, their mutual information is defined in terms of their PDF's *p(x), p(y)* and *p(x,y) as*:

$$I(X;Y) = \int \int p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{11}$$

Mutual information holds the following nice relationships with entropy:

$$I(X;Y) = I(Y;X) = \mathcal{H}(X) - \mathcal{H}(X|Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X) \tag{12}$$

where $\mathcal{H}(.|.)$ is the conditional entropy.

This application will utilize the mRMR or "minimum redundancy maximum relevance" (Peng et al, 2006) algorithm. This method computes MI using Shannon's entropy estimates. Principe (2010) states that this is computationally expensive compared to using Renyi's quadratic entropy estimate in equation (9), $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$, respectively.

Therefore, this research will use the following algorithm developed by C.T. Liu and B.G. Hu (2010), replacing Shannon's entropy with Renyi's.

Given the input data X, with Nsamples of Mvariables F=$\{f_i, i = 1 \ldots, M\}$, and the target variable *C,* then the complete algorithm is:

1) Set F="full input variable set" and S="empty set"

2) $\forall f_i \in F$, compute $I(C; f_i)$

3) Find the input variable $\hat{f}_i$ that maximizes $I(C; f_i)$, set $F \leftarrow F/\hat{f}_i$, $S \leftarrow \hat{f}_i$

4) Repeat steps a and b until desired number of m of input variables are selected

a) Compute MI between variables for all pairs of variables $I(f_i, f_s)$ with $f_i \in F$, $f_s \in S$

b) Choose the input variable $f_i \in F$ that maximizes

$$\text{Max}_{x_j \in F - S_{k-1}}[I(C; f_j) - \frac{1}{k-1}\Sigma_{x_i \in S_{m-1}} I(f_j; f_i)] \quad (13)$$

5) Output the subset S containing m significant input variables

3.4 Dataset

The “Hitters” dataset from the ISLR library in R was used. After cleaning for NA data values and removal of strictly categorical variables, this dataset contained 263 discrete observations ($N = 263$), 16 input variables ($m = 16$), with $C = \mathrm{Salary}$. The variable names were *'AtBat' , 'Hits', 'HmRun', 'Runs', 'RBI', 'Walks', 'Years', 'CAtBat', 'CHits', 'CHmRun', 'Cruns', 'CRBI', 'Cwalks', ‘PutOuts', 'Assists', and 'Errors'.*

3.5 Entropy Estimation for Variable Selection

A complete function in R was developed, which for each m input variable and C, computed σ_{opt} following equation (10) and then estimated $V_2(x)$ using the quadratic Gaussian kernel presented in equation (9). This function is presented in Appendix A. Finally $-log[\hat{V}_2(x)]$ was calculated for each variable, which is our estimate for quadratic Renyi entropy $\mathcal{H}_{(\mathrm{R},2)}(x)$ in equation (8).

Next, $I(C; f_i)$ was computed following the equations presented in (12) while substituting Renyi’s measure for $\mathcal{H}(.)$:

$$I(C; f_i) = \mathcal{H}(C) - \mathcal{H}(C|f_i). \tag{14}$$

This resulted in CAtBat to be the most informative significant input variable.

Finally, following equation (13), the order of variables, from most informative to least in terms of $C = \mathrm{Salary}$, was determined to be: *‘CAtBat’, ‘CHits’, ‘CRuns’, ‘CWalks’, ‘CRBI’, ‘PutOuts’, ‘AtBat’, ‘Assists’, ‘CHmRun’, ‘Hits’, ‘RBI’, ‘Runs’, ‘Walks’, ‘HmRun’, ‘Errors’, ‘Years’.*

At this point, the number of significant variables were not determined. Only the ordering was of concern here.

3.6 Methods of Comparison

The first method of comparison was the Lasso method for variable selection . The well-known advantage of the Lasso is that it it forces some coefficients to go towards exactly zero due to the use of an L_1 penalty, instead of an L_2 penalty. The reasoning behind selecting this method is that the Lasso is a variance-based measure (through the mean squared error (MSE)) to preform variable selection.

Preforming the Lasso on the "Hitters" data resulted in the following variables being selected: *'Hits', 'Walks', 'CHmRun', 'CRuns', 'CRBI', 'PutOuts',* with a test $\text{MSE} \approx 104841.34$ and $\lambda_{\text{lasso}} \approx 22$. The rest of the variables were forced to zero.

The second method of comparison was the use of Backward/Forward selection on the full "Hitters" data. In R, both methods can be ran simultaneously to ensure an accurate result. This is a popular method of variable selection that uses *Akaike Information Criterion* (AIC) to determine the best combination of variables using the information-theoretic concept of entropy maximization.

The results of the "Hitters" data model generalized linear negative-binomial Backward/Forward selection model in R were: *'AtBat', 'Hits', 'HmRun', 'Walks', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors'.*

These two methods are presented and compared to the Renyi Quadratic Entropy method of variable selection for very obvious, yet distinctive, reasons. The Lasso is a strong and popular variance-based method of variable selection, while Backward/Forward selection is another information-theoretic measure that is extremely popular. The comparison between these methods and the Renyi quadratic entropy estimation mRMR algorithm will be presented in the next section.

3.7 Results

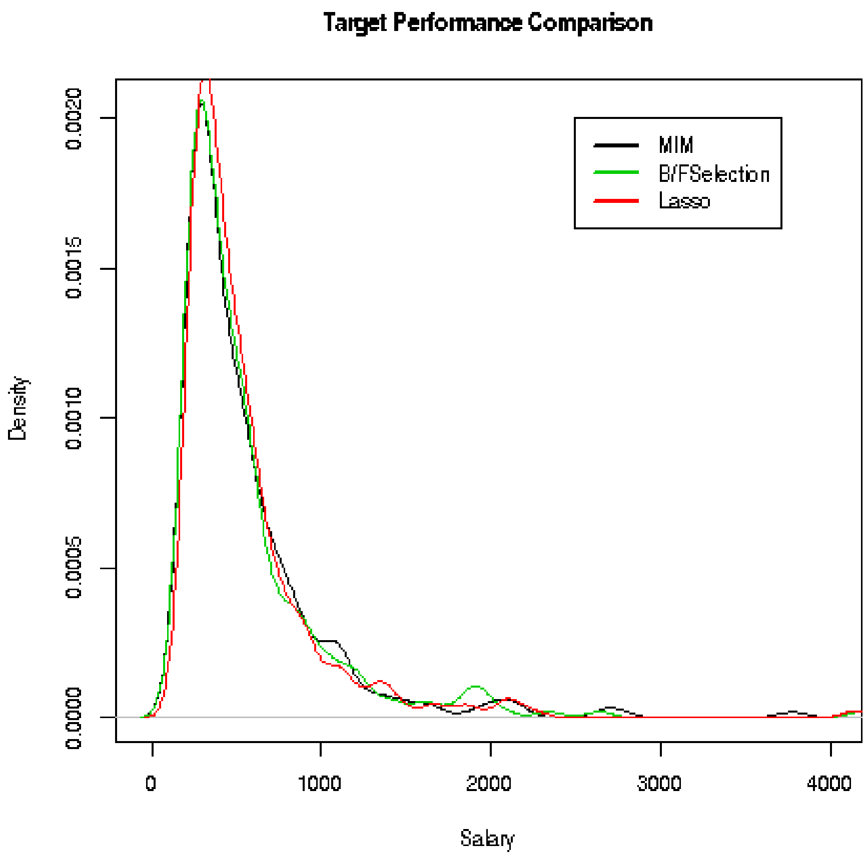
From the results in section 3.5, we obtained the ordering of the input variables that contained the most information with respect to the target variable, *Salary,* from the most informative ('*CAtBat'*) to the least informative (*'Years'*). With this in mind, multiple models were developed, each adding in the next ordered variable. These models were then compared to the model developed by the Lasso method which selected six input variables *('Hits', 'Walks', 'CHmRun', 'CRuns', 'CRBI', 'PutOuts'),* and the Backward/Forward model which selected nine input variables (*'AtBat', 'Hits', 'HmRun', 'Walks', 'CRuns', 'CWalks', 'PutOuts', 'Assists', 'Errors'*). Performance results are presented in Table 5.

Table 5: Method Statistics

Method	Mean	StdDev	Range	Median	IQR	3SD (Chebychev)	AIC
True Target	535.94	451.10	2392	425.00	560.00	(88.82,1305.9)	—
MIM	560.21	453.21	3574.72	407.95	378.82	(227.21, 1308.62)	3662.60
B/F Selection	560.67	465.36	4014.44	410.31	354.25	(220.36,1387.78)	3650.10
Lasso	551.88	426.69	3965.15	421.27	325.63	(232.61,1314.97)	3667.60

The mRMR model that performed the best, based on AIC and the neighborhood of correct prediction, included the first ten ordered input variables (*'CAtBat', 'CHits', 'CRuns', 'CWalks', 'CRBI', 'PutOuts', 'AtBat', 'Assists', 'CHmRun', 'Hits'*). We will refer to this model as the *Most Informative Model* (MIM) henceforth. The Backward/Forward selection model had the lowest AIC (3650.1), but had the worst prediction accuracy in regards to the target variable, especially in regards to extreme low (under 100) and high (over 1000) values. The individual comparisons of the model response densities are recorded in Figure 1.

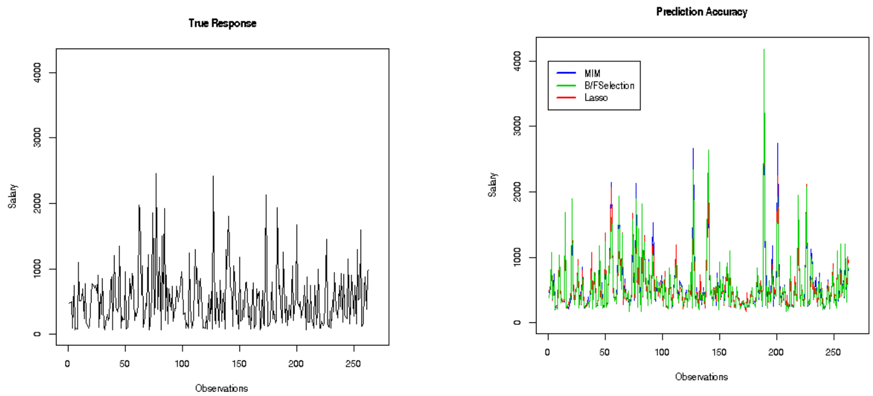
Figure 1: Density of Target Prediction Accuracy



The Lasso model had the highest AIC (3667.6). This model had better prediction accuracy than the Backward/Forward Selection Model, but also suffered with extreme values. The MIM fell in between these two models for AIC (3662.6), which was also close to the Full Model AIC (3661.5). This method was also more consistent with recognizing the extreme values than the other methods. This is easily recognized with the standard deviation and multiple range values (range, IQR, and density range based on 3SD using Chebychev's Rule) being closest to the true target density.

Comparison output based on individual ranked observations between the true target values and the model estimations is provided in Figure 2.

Figure 2: Comparison Between the True Target Response and Method Estimations



3.8 Discussion

The presented results demonstrate how the MIM developed through the Renyi quadratic entropy estimation mRMR algorithm was capable at preforming comparably well with more widely used methods for variable selection, and in some instances, better.

The MIM model was the only method that detected the input variables '*CAtBat'* and '*CHits'*, which this algorithm determined were the two most informative input variables. This implies that other variance or information theoretic based methods might inherently skip target-informative input variables by design. This model also appeared to narrow in on key components of the target variable, such as outliers, better than the other methods tested.

A couple downsides to this method was the complexity of the design and the selection of the most input variables to reach the conclusion, but the interpretation of this model is straightforward (the model contains the m input variables that contain the most relevant information in regards to the target variable).

There are further considerations and concerns regarding this research. While a sample size of 263 observations is considered large, in the scope of machine learning this would be considered small. Since this dataset was quite skewed, further research would be required to determine how the MIM's developed using the mRMR with Renyi Quadratic Entropy Estimation would compare to these presented methods and others using a much larger dataset. It appears a MIM based model was quicker at ackowledging the characteristics of the target variable, but replication would be required to conclude this issue.

Another issue that arose is whether comparing information-based selection models (Backward/Forward Selection model and the MIM) using information criteria (AIC) acceptable, such as was done in this research. A standard method of comparison between these types of models may need to be explored.

While there are some questions that arose in the process of comparison, for this easy application it appears that the MIM method is at least a relevant alternative to modern variable selection techniques. While the area of Information Theoretic Learning is a still young field, this application provides a small insight into the implications the area has on many disciplines, including statistics.

4. CONCLUSION

This research explored the information-theoretic measure of the Renyi Spectra of Information from both a parametric and non-parametric perspective. In section 2, Renyi's measure was computed for twenty discrete and continuous univariate parametric families, and the behavior of their entropy and variance functions were evaluated for distinct similarities and disagreements to assist in relating the two uncertainty measures. In section 3, a variable selection algorithm that incorporates the non-parametric kernel estimation of Renyi's Quadratic Entropy was presented and compared to popular variance-based (the Lasso) and information-theoretic (Backward/Forward selection) methods for variable selection, with respectable results.

The goal of this paper is to discuss Renyi's entropy and other information-theoretic measures from a statistical perspective. Section 2 provides comprehensive lists and discussions that can be applicable across many different disciplines, while section 3 provides an alternative approach to machine learning (Information Theoretic Learning) with an application in the important and much discussed area of variable selection.

Further extensions of this research include exploration into the behavior of Renyi's measure and variance for multivariate distributions and comparisons of other variable selection methods using the Renyi quadratic entropy estimation mRMR algoithm.

REFERENCES

Cover T.M. & Thomas J.A., Elements of Information Theory, 2nd edition. NewYork: John Wiley, 2005.

Ebrahimi,N., Maasoumi, E. & Soofi, E.S. (1999). Ordering univariate distributions by entropy and variance. *J. Econometrics, 90*, 317-336.

Ebrahimi, N., Soofi, E. S., & Soyer, R. (2010). Information measures in perspective. *International Statistical Review, 78*(3), 383-412.

James, G., Witten, D., Hastie, T. & Tibshirani, R., An Introduction to Statistical Learning with Applications in R. New York, NY: Springer Publishing, 2013.

Liu, C.T. & Hu, B.G. (2010). Renyi's-entropy-based approach for selection the significant input variable for the ecological data. In *International Congress on environmental modelling and software.,* 487.

Nadarajah, S. & Zografos, K. (2003). Formulas for Renyi information and related measures for univariate distribution. *Inform. Sci., 155*, 119-138.

Peng H., Long, F. and Ding C. (2005) Feature selection based on MI: Criteria of max-dependency, max-relevance and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, *27*(8), 1226–1238.

Principe, J.C., Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives. New York, NY: Springer Publishing, 2010.

Renyi, A. (1961). On measures of entropy and information. In *Proc. Fourth Berkeley Symp., 1*, 547-561. Berkeley: UC Press.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379-423.

Song, K.S. (2001) Renyi information, loglikelihood, and a intrinsic distribution measure. *Journal of Statistical Planning and Inference, 93*, 51-69.

APPENDIX A
RENYI'S QUADRATIC ENTROPY KERNEL DENSITY ESTIMATION R CODE

```
renEst <- function(x,i){

   sdopt <- function(x,i){
      for(i in 1:dim(x)[2]){
   i = as.character(names(x))
   sdopt <- (diag(sqrt(abs(var(x[i]))))*(((4/dim(x)[1])/(2*dim(x)[2] + 1))^(1/(dim(x)[2] +
4)))))/sqrt(2)
      }
   return(sdopt[i])
   }

   sd <- sdopt(x,i)

   Est <- function(x,i){

   efx <- try(density(as.numeric(unlist(x[i])), kernel = "gaussian", bw = as.numeric(sd)[i],
         adjust=1), silent=TRUE)
   return(efx)
   }
   results <- Est(x,i)
   return(results)
   }
```

APPENDIX B
ORDERING EXAMPLE R CODE

Generalized-normal:

```
alp = 1
bet = 5
lambda = seq(1,2,length=10)
x=2
y = (1/(1-lambda))*(-lambda*log(2*bet*gamma(alp/2))) +
((-alp*lambda + lambda - 1)/2)*log((bet*lambda)/2) +
log(pgamma((lambda*x^(2))/bet^2,((alp+1)*lambda + 1)/2, lower =
TRUE)*gamma(((alp+1)*lambda + 1)/2))

alp
bet
y
```