"Mircea cel Batran" Naval Academy Scientific Bulletin, Volume XVIII – 2015 – Issue 2
Published by "Mircea cel Batran" Naval Academy Press, Constanta, Romania // The journal is indexed in:
PROQUEST SciTech Journals, PROQUEST Engineering Journals, PROQUEST Illustrata: Technology, PROQUEST Technology Journals, PROQUEST Military Collection PROQUEST Advanced Technologies & Aerospace

ON ENTROPIC MEASURES FOR LOGISTIC SEMIPARAMETRIC REGRESSION MODELS WITH BINARY RESPONSE

George-Andrei PADUREANU[1]
Monica DUMITRESCU[2]
[1] Ph.D. student University of Bucharest / Doctoral School in Mathematics, Bucharest, Romania
[2] Prof. Dr. University of Bucharest / Faculty of Mathematics and Computer Science, Bucharest, Romania

***Abstract**: The paper introduces two measures for the informational properties of regression models which deal with a random vector (response, covariates) and are based on the assumption of the existence of an intrinsic relationship between covariates (causes) and response (effect). We define these measures in terms of conditional Shannon entropy and conditional α-Renyi entropy. The regression models we address are logistic semiparametric regression models with binary response (LSpRModelsBR) and with two exogenous covariates. Conditional entropies are defined and calculated for discrete, binary covariates and for exponential distributed covariates, the issue of nonparametric estimation of the conditional quadratic Renyi entropy is discussed and we report the results of a simulation study. Based on their properties and on our simulation results, we conclude that these conditional entropies are able to measure the intensity of the connection "response ~covariates" within a regression model. Therefore, we can identify a new goodness-of-fit index for regression models, as well as a new quantitative criterion for statistical modelling:*
"The larger conditional entropy H(response | covariates), the better fitted the regression model response ~covariates".
***Keywords**: conditional entropy, kernel estimators, semiparametric logistic regression.*

INTRODUCTION

Linear generalized semiparametric regression models become very popular tools for statistical modelling, as they emphasize a remarkable flexibility and allow application of efficient mathematical techniques for data analysis. Understanding their deeper properties can be extremely useful for all stages, modelling, estimation, statistical testing. Therefore, finding the informational properties of these regression models would be an issue of high interest, both theoretical and practical.

In any regression model, one deals with a random vector (response, covariates) and the existence of an intrinsic relationship between covariates (causes) and response (effect). This relationship is essential for modelling and its informational dimension could be expressed in terms of a conditional entropy H (response | covariates): It is our task in this paper to define, calculate and estimate such entropic measures, and we present the results only for one class of models, logistic semiparametric regression models with binary response (LSpRModelsBR) and with two exogenous covariates.

We discuss two approaches on conditional entropy: the Shannon conditional entropy (which is clearly defined and well studied) [10] and the α-Renyi conditional entropy [11] (for which there is no unanimously accepted definition). We substantiate our choice for the Jizba & Arimitsu [3] version of the α-Renyi conditional entropy. The expressions of these two entropic measures are presented for two cases, a regression model with discrete covariates and another one, with continuous covariates. Also, we discuss the issue of nonparametric estimation for these conditional entropies. The paper closes with a simulation study.

A. The Shannon and conditional Shannon entropy

The well known definitions (Shannon, 1948) [10] of these entropic measures are:

$$H^S(X) = E_X\left(-\log P_X(x)\right) = -\sum_x P_X(x)\log P_X(x),$$

$$\begin{aligned} H^S(X \mid Y) &= E_y\left(H^S(X \mid Y = y)\right) \\ &= \sum_y P_Y(y)\left(-\sum_x P_{X|Y}(x \mid y)\log P_{X|Y}(x \mid y)\right). \end{aligned} \qquad (1)$$

We mention some of the most important properties, which allow the interpretation of the Shannon entropy as a measure of uncertainty associated with the (discrete) probability distributions of the involved random variables:

- Positivity:

$$0 \le H^S(X) \le \log\left(card\left(X(\Omega)\right)\right)$$

- Subaditivity
$$H^S\left((X,Y)\right) \le H^S(X) + H^S(Y)$$
- Additivity of independent systems:
$$H^S((X,Y)) = H^S(X) + H^S(Y)$$
$$\Leftrightarrow X, Y \ independent$$
- Monotonicity
$$H^S(X \mid Y) \le H^S(X)$$
- STRONG CHAIN RULE
$$H^S(X \mid Y) = H^S((X,Y)) - H^S(Y)$$
- CHAIN RULE
$$H^S(X \mid Y) \ge H^S((X,Y)) - \log(card(Y(\Omega)))$$
$$\ge H^S(X) - \log(card(Y(\Omega))).$$
- AIE (Additional information Increases Entropy), which implies that the entropy of X increases if some information is added.

TABLE I.
DEFINITIONS OF CONDITIONAL RENYI ENTROPY

	$H_\alpha^{R;1}(X \mid Y)$	$H_\alpha^{R;2}(X \mid Y)$	$H_\alpha^{R;3}(X \mid Y)$
non-negativity	Yes	Yes	Yes
Chain rule	No	Yes	No
Strong chain rule	No	Yes	Yes
conditioned AIE	Yes	Yes	Yes

The corresponding definitions for continuous random variables are:
$$H^S(X) = E_X\left(-\log f_X(x)\right)$$
$$= -\int f_X(x) \ln f_X(x) dx,$$
$$H^S(X \mid Y) = E_y\left(H^S(X \mid Y = y)\right)$$
$$= \int f_Y(y)\left(-\int f_{X|Y}(x \mid y) \ln f_{X|Y}(x \mid y) dx\right) dy.$$

B. The α-Renyi entropy

In 1961, Renyi generalized the Shannon entropy by modifying one of its axioms characterizing the averaging of information. Renyi's entropies keep Shannon additivity property of independent systems.

$$H_\alpha^R(X) = \frac{1}{1-\alpha} \log\left(\sum_x \left(P_X(x)\right)^\alpha\right)$$
$$= -\log\left(E_X\left(P_X(x)\right)^{\alpha-1}\right)^{1/\alpha-1} \quad \alpha \ge 0, \quad \alpha \ne 1$$

Here are some properties of interest for our work:

- H_α^R is concave and monotonically decreasing in α:
$$H_\alpha^R(X) \le H_\beta^R(X) \quad , \quad \forall \quad \alpha \ge \beta$$
- H_α^R is consistent with Shannon's entropy:
$$\lim_{\alpha \to 1} H_\alpha^R(X) \le H^S(X)$$
- H_α^R is consistent with collision entropy:
$$H_2^R(X) = -\log\left(\sum_z \left(P_X(x)\right)^2\right) \text{ (Collision entropy)}$$
- Additivity of independent systems:
$$H_\alpha^R((X,Y)) = H_\alpha^R(X) + H_\alpha^R(Y), \ for \quad X, Y \quad independent \quad .$$

The corresponding definition for continuous random variables is:
$$H_\alpha^R(X) = \frac{1}{1-\alpha} \log\left(\int (f_X(x))^\alpha dx\right)$$
$$= -\log\left(E_X\left(f_X(x)\right)^{\alpha-1}\right)^{1/\alpha-1} \quad \alpha \ge 0, \quad \alpha \ne 1.$$

C. The conditional Renyi entropy

There is no commonly accepted definition of conditional Renyi entropy. We present three proposals for this entropic measure:

- the Cachin definition (1997) [2]

$$H_\alpha^{R;1}(X \mid Y) = \sum_y P_Y(y) H_\alpha^R(X \mid Y = y),$$

- the Jizba & Arimitsu definition (2004) **[3]**, as well as the Golshani, Pasha &Yari definition (2009) **[4]**

$$H_\alpha^{R;2}(X \mid Y) = H_\alpha^R(X,Y) - H_\alpha^R(Y),$$

- the Renner & Wolf definition (2004) **[5]**

$$H_\alpha^{R;3}(X \mid Y) = \min_y H_\alpha^R(X \mid Y = y).$$

Teixeira, Matos & Antunes (2012) **[7]** compare these three definitions and their findings are presented in the following Table I.

Based on these properties, as well as on the existence of an intrinsic relationship between covariates (causes) and response (effect) in a regression model, we adopt the Jizba & Arimitsu definition **[3]** for the conditional α-Renyi entropy, which we simply denote $H_\alpha^R(X \mid Y)$:

$$H_\alpha^R(X \mid Y) = H_\alpha^R(X,Y) - H_\alpha^R(Y) \quad \textbf{[7]}$$

As already mentioned, our task in this paper is to define, calculate and estimate these entropic measures for one class of models, logistic semiparametric regression models with binary response (LSpRModelsBR) and with two exogenous covariates.

The paper is organized as follows: in *Section 2* we calculate the conditional Shannon entropy and the conditional α-Renyi entropy for **LSpRModelsBR**, in *Section 3* we construct a nonparametric kernel estimator for the introduced quadratic conditional Renyi entropy, and in *Section 4* we report the results of a simulation study.

CONDITIONAL ENTROPIES FOR LSPRMODELSBR

Logistic semiparametric regression models with binary response are defined by the following items:

- There are two (vector) covariates X = ($X_1, \dots, X_r$) and U =($U_1, \dots, U_{d-r}$) , which are independent (exogenous);
- The distributions of covariates are either discrete (with probabilities ($P_{\mathbf{X}}(\mathbf{x})$, $\mathbf{x}$) , ($P_{\mathbf{U}}(\mathbf{u})$, $\mathbf{u}$)), or continuous (with densities $f_{\mathbf{X}}(\mathbf{x})$, $f_{\mathbf{U}}(\mathbf{u})$);
- The response *Y* is a binary variable, with a Bernoulli conditional distribution $Y \mid (\mathbf{x}, \mathbf{u}) \sim B(1, \pi(\mathbf{x}, \mathbf{u}))$,

$$\pi_{(\mathbf{x},\mathbf{u})} = \frac{\exp(\delta + \beta \cdot \mathbf{x} + g(\mathbf{u}))}{1 + \exp(\delta + \beta \cdot \mathbf{x} + g(\mathbf{u}))}$$

$$\beta = (\beta_1, \dots, \beta_r)^t$$

Function g should be a smooth one, satisfying continuity conditions on g and on its first two derivatives.

The conditional Shannon entropy associated with this model is

$$H^S(Y \mid (\mathbf{X},\mathbf{U})) = E_{(\mathbf{X},\mathbf{U})}\left(H^S(Y \mid (\mathbf{X},\mathbf{U}) = (\mathbf{x},\mathbf{u}))\right), \tag{3}$$

Also, the conditional Renyi entropy for this regression model is:

$$H_\alpha^R(Y \mid (\mathbf{X},\mathbf{U})) = H_\alpha^R(Y,\mathbf{X},\mathbf{U}) - H_\alpha^R(\mathbf{X},\mathbf{U})$$

Its non-negativity, additivity of independent systems and the fact that it satisfies chain rules makes it a very good candidate for a measure of the intensity of the connection between the covariates (X, U) and the response Y.

D. The model with d = 2, r = 1 and binary covariates

Let us denote by M(1) a logistic semiparametric regression model with binary response and with two exogenous binary covariates,

$$\mathcal{M}(1) \ : \begin{cases} (Y,(X,U)) = (response, covariates) \\ X,U \text{ independent} \\ X \sim B(1,\tau), \quad U \sim B(1,\nu), \quad 0 < \tau,\nu < 1 \\ Y \mid (x,u) \sim B(1,\pi(x,u)) \end{cases}$$

$$\pi_{(x,u)} = \frac{\exp(\delta + \beta \cdot x + g(u))}{1 + \exp(\delta + \beta \cdot x + g(u))}$$

Proposition 1

The conditional Shannon entropy associated with M (1) is

$$H^S(Y\,|\,(X,U)) = \\ = -\sum_{x,u=0,1}\tau^x(1-\tau)^{1-x}\nu^u(1-\nu)^{1-u}\cdot\left(\pi_{(x,u)}\log\pi_{(x,u)}+\left(1-\pi_{(x,u)}\right)\log\left(1-\pi_{(x,u)}\right)\right). \quad (5)$$

The conditional Renyi entropy associated with M (1) is

$$H_\alpha^R(Y\,|\,(X,U)) = \\ = \frac{1}{1-\alpha}\log\left(\frac{\sum_{y,x,u=0,1}\left(\tau^x(1-\tau)^{1-x}\nu^u(1-\nu)^{1-u}\cdot\left(\pi_{(x,u)}\right)^y\left(1-\pi_{(x,u)}\right)^{1-y}\right)^\alpha}{\left(\tau^\alpha+(1-\tau)^\alpha\right)\cdot\left(\nu^\alpha+(1-\nu)^\alpha\right)}\right) \quad (6)$$

Proof:

The conditional Shannon entropy is given by the expression

$$H^S(Y\,|\,(X,U)) = \sum_{x,u}P_{(X,U)}(x,u)\left(-\sum_y P_{Y|(X,U)}(y\,|\,(x,u))\log P_{Y|(X,U)}(y\,|\,(x,u))\right)$$

By direct calculation we obtain the form (5) of this entropy.
The conditional α-Renyi entropy is given by the expression

$$H_\alpha^R(Y\,|\,(X,U)) = H_\alpha^R(Y,X,U) - H_\alpha^R(X,U),$$

where

$$H_\alpha^R(X,U) = \frac{1}{1-\alpha}\log\left(\sum_{x,u=0,1}\left(\tau^x(1-\tau)^{1-x}\nu^u(1-\nu)^{1-u}\right)^\alpha\right) \\ = \frac{1}{1-\alpha}\log\left(\left(\tau^\alpha+(1-\tau)^\alpha\right)\cdot\left(\nu^\alpha+(1-\nu)^\alpha\right)\right)$$

By direct calculation we obtain the expression (6) for the conditional Renyi entropy associated with M (1).

E. The model with d = 2; r = 1 and Exponential distributed covariates

Let us denote by M (2) a logistic semiparametric regression model with binary response and with two exogenous, Exponential distributed covariates,

$$\mathcal{M}(2)\;:\begin{cases} (Y,(X,U)) = (response, covariates) \\ X,U \text{ independent} \\ X\sim Exp(\lambda_1),\quad U\sim Exp(\lambda_2),\quad \lambda_1,\lambda_2>0 \\ Y\,|\,(x,u)\sim B(1,\pi(x,u)) \end{cases} \quad \pi_{(x,u)} = \frac{\exp(\delta+\beta\cdot x+g(u))}{1+\exp(\delta+\beta\cdot x+g(u))}$$

Proposition 2

The conditional Shannon entropy associated with M (2) is

$$H^S(Y\,|\,(X,U)) = \\ = -\lambda_1\lambda_2\int_0^\infty\int_0^\infty\exp(-\lambda_1 x-\lambda_2 u)\left(\pi_{(x,u)}\log\pi_{(x,u)}+\left(1-\pi_{(x,u)}\right)\log\left(1-\pi_{(x,u)}\right)\right)dxdu \quad (7)$$

The conditional Renyi entropy associated with M (2) is

$$H_\alpha^R(Y\,|\,(X,U)) = \\ = \frac{\log\lambda_1+\log\lambda_2+2\log\alpha}{1-\alpha} + \\ + \frac{1}{1-\alpha}\log(\int_0^\infty\int_0^\infty e^{-\alpha\lambda_1\cdot x}\cdot e^{-\alpha\lambda_2\cdot u}\left(\pi_{(x,u)}\right)^\alpha dxdu + \\ + \int_0^\infty\int_0^\infty e^{-\alpha\lambda_1\cdot x}\cdot e^{-\alpha\lambda_2\cdot u}\left(1-\pi_{(x,u)}\right)^\alpha dxdu) \quad (8)$$

Proof:

The conditional Shannon entropy is

$$H^S(Y\,|\,(X,U)) = \int_0^\infty\int_0^\infty f_X(x)f_U(u)\cdot \\ \cdot\left(-\sum_{y=0,1}P_{Y|(X,U)}(y\,|\,(x,u))\log P_{Y|(X,U)}(y\,|\,(x,u))\right)dxdu$$

and, by direct calculation we obtain the expression (7)

The conditional α-Renyi entropy is given by the relation

$$H_\alpha^R(Y \mid (X,U)) = H_\alpha^R(Y,X,U) - H_\alpha^R(X,U) \quad (9)$$

where

$$H_\alpha^R(X,U) = H_\alpha^R(X) + H_\alpha^R(U) =$$
$$= \frac{1}{1-\alpha}\left(\log\left(\int_0^\infty \left(\lambda_1 e^{-\lambda_1 \cdot x}\right)^\alpha dx\right) + \log\left(\int_0^\infty \left(\lambda_2 e^{-\lambda_2 \cdot u}\right)^\alpha du\right)\right)$$
$$= \frac{1}{1-\alpha}\left(\log\left(\frac{\lambda_1^{\alpha-1}}{\alpha}\right) + \log\left(\frac{\lambda_2^{\alpha-1}}{\alpha}\right)\right)$$
$$H_\alpha^R(X,U) = \frac{2\log\alpha}{\alpha-1} - (\log(\lambda_1) + \log(\lambda_2)) \quad (0)$$

In order to calculate $H_\alpha^R(Y,X,U)$ we need to define a "density" for the vector (*Y, X, U*), which has a discrete component and two continuous components.
Notice that the following relation is verified:

$$\sum_{y=0,1}\left(\int_0^\infty\int_0^\infty f_X(x)\cdot f_U(u)\cdot\left(\pi_{(x,u)}\right)^y\cdot\left(1-\pi_{(x,u)}\right)^{1-y} dxdu\right) = 1$$

Hence, we can define the "density" of (*Y, X, U*) by the expression

$$f_{(Y,X,U)}(y,x,u) = f_X(x)\cdot f_U(u)\cdot\left(\pi_{(x,u)}\right)^y\cdot\left(1-\pi_{(x,u)}\right)^{1-y}$$
$$f_{(Y,X,U)}(y,x,u) = \begin{cases}\lambda_1 e^{-\lambda_1\cdot x}\cdot\lambda_2 e^{-\lambda_2\cdot u}\cdot\pi_{(x,u)}, & y=1\\ \lambda_1 e^{-\lambda_1\cdot x}\cdot\lambda_2 e^{-\lambda_2\cdot u}\cdot\left(1-\pi_{(x,u)}\right), & y=0\end{cases}\quad, x,u\ge 0 \quad (11)$$

Then, the α-Renyi entropy $H_\alpha^R(Y,X,U)$ is

$$H_\alpha^R(Y,X,U)$$
$$= \frac{1}{1-\alpha}\log\left(\sum_{y=0,1}\int_0^\infty\int_0^\infty\left(\lambda_1 e^{-\lambda_1\cdot x}\right)^\alpha\cdot\left(\lambda_2 e^{-\lambda_2\cdot u}\right)^\alpha\cdot\left(\pi_{(x,u)}\right)^{\alpha y}\cdot\left(1-\pi_{(x,u)}\right)^{\alpha(1-y)} dxdu\right)$$

$$H_\alpha^R(Y,X,U)$$
$$= \frac{1}{1-\alpha}\log\Big(\int_0^\infty\int_0^\infty\left(\lambda_1 e^{-\lambda_1\cdot x}\right)^\alpha\cdot\left(\lambda_2 e^{-\lambda_2\cdot u}\right)^\alpha\cdot\left(\pi_{(x,u)}\right)^\alpha dxdu +$$
$$+\int_0^\infty\int_0^\infty\left(\lambda_1 e^{-\lambda_1\cdot x}\right)^\alpha\cdot\left(\lambda_2 e^{-\lambda_2\cdot u}\right)^\alpha\cdot\left(1-\pi_{(x,u)}\right)^\alpha dxdu\Big)$$

$$H_\alpha^R(Y,X,U) \quad (12)$$
$$= \frac{\alpha(\log\lambda_1 + \log\lambda_2)}{1-\alpha} + \frac{1}{1-\alpha}\log\Big(\int_0^\infty\int_0^\infty\left(\lambda_1 e^{-\lambda_1\cdot x}\right)^\alpha\cdot\left(\lambda_2 e^{-\lambda_2\cdot u}\right)^\alpha\cdot\left(\pi_{(x,u)}\right)^\alpha dxdu +$$
$$+\int_0^\infty\int_0^\infty\left(\lambda_1 e^{-\lambda_1\cdot x}\right)^\alpha\cdot\left(\lambda_2 e^{-\lambda_2\cdot u}\right)^\alpha\cdot\left(1-\pi_{(x,u)}\right)^\alpha dxdu\Big)$$

Replacing the expressions (10) and (12) into (9) we get the form (8) of the conditional α-Renyi entropy.
A KERNEL ESTIMATOR OF THE QUADRATIC CONDITIONAL RENYI ENTROPY
The estimation of Shannon's or Renyi's entropy directly from data would follow the route: **data → pdf estimation →integral estimation**. By this strategy, one obtains the "resubstitution" estimate.
Notice that entropy is a scalar, but as an intermediate step one would have to estimate a function (the pdf). Therefore, bypassing the stage of pdf estimation as a function would be very convenient.

F. General results

1) Kernel estimators for the probability density

of a random variable have been introduced by Parzen (1962) [12]:

$$\hat{f}_Z(z) = \frac{1}{N\cdot\sigma}\sum_{i=1}^N K\left(\frac{z-z_i}{\sigma}\right)$$

where σ is a bandwidth parameter and the kernel function satisfies the conditions

$$K(z) \geq 0$$
$$\int_R K(z)dz = 1$$
$$\lim_{z\to\infty} |z \cdot K(z)| = 0$$

Usually, kernels are symmetrical, normalized, unimodal, continuous and differentiable functions. The most used function is the *Gaussian kernel with standard deviation σ,*

$$G_\sigma(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right)$$

2) *Kernel estimators for a discrete distribution with integer, positive values*

Kernel estimators for a discrete distribution with integer, positive values have been discussed by Rajagopalan & Lall (1995) **[8]** who have considered weighted linear combinations of relative frequencies in the sample:

$$\hat{P}_Z(i) = \sum_{j=1}^{k_{\max}} K(z_j)\tilde{P}_j$$
$$i, j, h \in N,$$
$$z_j = \frac{i-j}{h}$$

where *K (z)* is a kernel function, and $\tilde{P}_j$ is the relative frequency of *j* in the sample.

Wang and Van Ryzin (1981) **[9]** have proposed the use of a Geometric kernel for nonparametric estimation,

$$K_h(i, j) = \begin{cases} \frac{1}{2}h(1-h)^{|i-j|} & , \quad |i-j| \geq 1 \\ 1-h & , \quad i = j \end{cases}$$
$$h \in (0,1)$$

Notice that, in this discrete case, estimation of the probabilities is reached by direct calculation, as just a finite number of relative frequencies are involved in the formulas.

3) *Resubstitution kernel estimator of the conditional Shannon entropy*

Resubstitution kernel estimator of the conditional Shannon entropy from a sample ((y_1, x_1, u_1) , ..., (y_N, x_N, u_N)) has the form

$$\hat{H}^S(Y \mid (X,U)) = -\frac{1}{N}\sum_{i=1}^{N} \log \hat{f}_{(Y|(X,U))}(y_i \mid (x_i, u_i)) \qquad ,$$

where $\hat{f}_{(Y|(X,U))}$ is a kernel estimator of the conditional density. More details can be found in Beirlant et al (1997) **[1]**.

G. *Kernel estimators of the quadratic conditional Renyi entropy*

Xu and Erdogmuns (2010) **[6]** have introduced an estimator of the quadratic Renyi entropy, bypassing the explicit need to estimate the pdf $f_Z(z)$, as only $E_Z(f_Z(z))$ is needed, which is a scalar.

$$H_2^R(Z) = -\log\left(\int (f_Z(z))^2 dz\right) = -\log(E_Z(f_Z(z)))$$

The Xu and Erdogmuns estimator **[6]** constructed from a sample (z_1, ..., z_N) is

$$\hat{H}_2^R(Z) = -\log \int_{-\infty}^{\infty} \left(\frac{1}{N}\sum_{i=1}^{N} G_\sigma(z - z_i)\right)^2 dz$$
$$= -\log \frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N} G_{\sigma\sqrt{2}}(z_j - z_i)$$

where $G_\sigma(\cdot)$ is the Gaussian kernel with standard deviation σ.

The estimation of the quadratic conditional Renyi entropy for **LSpRModelsBR** reduces to the estimation of the quadratic Renyi entropy for the covariates and for the vector (response, covariates),

$$\hat{H}_2^R(Y \mid (X,U)) = \hat{H}_2^R(Y, X, U) - \hat{H}_2^R(X, U) \quad (13)$$

Therefore, we need to extend the Xu and Erdogmuns construction **[6]** to (bidimensional and tridimensional) random vectors. The construction of kernel estimators is performed on the basis of a set of independent observations

$$\{(y_{ij},(x_i,u_i) \quad , \quad j=1,...,M) \quad , \quad i=1,...,N\}$$

1) *Logistic semiparametric regression model with two exogenous discrete covariates.*

For the covariate *X*, we have the estimator

$$\hat{P}_X(x)=\sum_{x_j} K_h(x,x_j)\tilde{p}_j$$

where $\tilde{p}_j$ is the frequency of x_j in the sample and $K_h(x,x_j)$ is the Geometric kernel. Similar, for U we have the estimator

$$\hat{P}_U(u)=\sum_{u_j} K_h(u,u_j)\tilde{q}_j$$

where e $\tilde{q}_j$ is the frequency of u_j in the sample and $K_h(u,u_j)$ is the Geometric kernel.

Since X and U are independent, we can use the relation

$$\hat{H}_2^R(X,U)=\hat{H}_2^R(X)+\hat{H}_2^R(U)$$

and the estimator $\hat{H}_2^R(X,U)$ is given by the expression

$$\hat{H}_2^R(X,U)=-\log\sum_{x=0}^{x_{\max}}\left(\sum_{x_j} K_h(x,x_j)\tilde{p}_j\right)^2 -\log\sum_{u=0}^{u_{\max}}\left(\sum_{u_j} K_h(u,u_j)\tilde{q}_j\right)^2 \qquad (4)$$

Kernel estimators of conditional probabilities are

$$\hat{P}_Y(y\mid(x_t,u_t))=\sum_{y_j} K_h(y,y_j)\tilde{w}_{j|(x_t,u_t)} \ ,$$

where $\tilde{w}_{j|(x_t,u_t)}$ is the frequency of y_j in the subsample (y_{tk}, (x_t; u_t) , k = 1,…,M) and $K_h(y,y_j)$ is a Geometric kernel.

Then, the estimator $\hat{H}_2^R(Y,X,U)$ is given by the expression

$$\hat{H}_2^R(Y,X,U) = -\log(\sum_{x,u}\left(\sum_{x_j} K_h(x,x_j)\tilde{p}_j\right)^2\left(\sum_{u_j} K_h(u,u_j)\tilde{q}_j\right)^2 \cdot \cdot\sum_{y}\left(\sum_{y_j} K_h(y,y_j)\tilde{w}_{j|(x_t,u_t)}\right)^2) \qquad (15)$$

Proposition 3

For a logistic semiparametric regression model with two exogenous discrete covariates, the kernel estimator $\hat{H}_2^R(Y\mid(X,U))$ is obtained by replacing relations (14) and (15) into relation (13).

In our simulation study, we apply this result for the regression model M (1).

2) *Logistic semiparametric regression model with two exogenous continuous covariates*

Now, let us consider a logistic semiparametric regression model with two exogenous continuous covariates. Then,

$$\hat{f}_X(x)=\frac{1}{N}\sum_{i=1}^{N} G_{\sigma,X}(x-x_i)$$

$$\hat{f}_U(u)=\frac{1}{N}\sum_{i=1}^{N} G_{\sigma,U}(u-u_i)$$

$G_\sigma(\cdot)$ is the Gaussian kernel with standard deviation σ .

The kernel estimator of $\hat{H}_2^R(X,U)$is obtained by direct calculation:

$$\begin{aligned}
&\hat{H}_2^R(X,U)= \\
&=-\log\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{N}\sum_{i=1}^{N}G_{\sigma,X}(x-x_i)\right)^2\left(\frac{1}{N}\sum_{i=1}^{N}G_{\sigma,U}(u-u_i)\right)^2 dxdu \\
&=\hat{H}_2^R(X)+\hat{H}_2^R(U)
\end{aligned}$$

with

$$\begin{aligned}
\hat{H}_2^R(X)&=-\log\frac{1}{N^2}\int_{-\infty}^{\infty}\sum_{j=1}^{N}\sum_{i=1}^{N}G_{\sigma,X}(x-x_i)G_{\sigma,X}(x-x_j)dx \\
&=-\log\frac{1}{N^2}\int_{-\infty}^{\infty}\sum_{j=1}^{N}\sum_{i=1}^{N}G_{\sigma\sqrt{2},X}(x_j-x_i)dx
\end{aligned}$$

(16) (16)

$$\begin{aligned}
\hat{H}_2^R(U)&=-\log\frac{1}{N^2}\int_{-\infty}^{\infty}\sum_{j=1}^{N}\sum_{i=1}^{N}G_{\sigma,U}(u-u_i)G_{\sigma,U}(u-u_j)du \\
&=-\log\frac{1}{N^2}\int_{-\infty}^{\infty}\sum_{j=1}^{N}\sum_{i=1}^{N}G_{\sigma\sqrt{2},U}(u_j-u_i)du
\end{aligned}$$

(17) (17)

On the other hand, we have

$$\hat{P}_Y(y\,|\,(x_t,u_t))=\sum_{y_j}K_h(y,y_j)\tilde{w}_{j|(x_t,u_t)}$$

Hence

$$\begin{aligned}
&\hat{H}_2^R(Y,X,U) \\
&=-\log(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{1}{N}\sum_{i=1}^{N}G_{\sigma,X}(x-x_i)\right)^2\cdot\left(\frac{1}{N}\sum_{i=1}^{N}G_{\sigma,U}(u-u_i)\right)^2\cdot \\
&\cdot\sum_{y}\left(\sum_{y_j}K_h(y,y_j)\tilde{w}_{j|(x_t,u_t)}\right)dxdu
\end{aligned}$$

(18)

Proposition 4

For a logistic semiparametric regression model with two exogenous continuous covariates, the kernel estimator $\hat{H}_2^R(Y,X,U)$ is obtained by replacing relations (16), (17) and (18) into relation (13).

In our simulation study, we apply this result for the regression model M(2) .

A SIMULATION STUDY

In our simulation study, we apply Proposition 3 to the regression model M(1) and Proposition 4 to the regression model M(2)

H. Model M(1)

Simulation generates N subsamples of shape (y_{tk}, $(x_t; u_t)$, k = 1,…,M) , on which are calculated estimators of conditional entropies, by applying in the following way:

Step 1: generate X~B(1,τ)

Step 2: generate U~B(1,μ)

Step 3: for each generated pair from Steps 1 and 2 generate M occurrences of Y as described in Section II.A.

Step 4: Calculate estimators of H(Y,X,U) and H(X,U)

Step 5: Apply Proposition 3 and calculate kernel estimators for conditional entropies

Table II shows parameters of this simulation.

TABLE II.
SIMULATION PARAMETERS FOR M(1)

	M	N	τ	μ	h
Range	10-1000	10-1000	0-1 by 0.2	0-1 by 0.2	0-1 by 0.2
	δ	β			
Range	1-5	1-5			

Figure1 shows relations between estimator of Conditional Entropy and Kernel estimator of H(Y,X,U) Figure 2 represents entropies as expressed in Proposition 3. Conditionality can be observed in both graphs. Calculated RMSE on simulated data, on different simulations parameters is retrieved in Figure3.

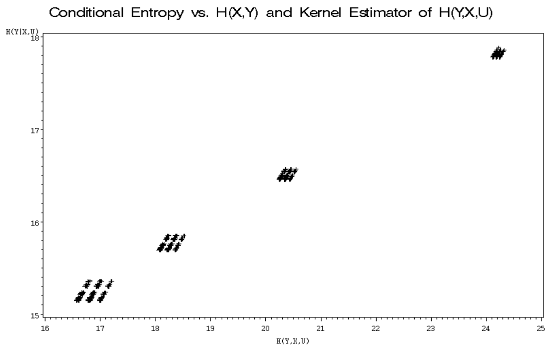

Figure 1.

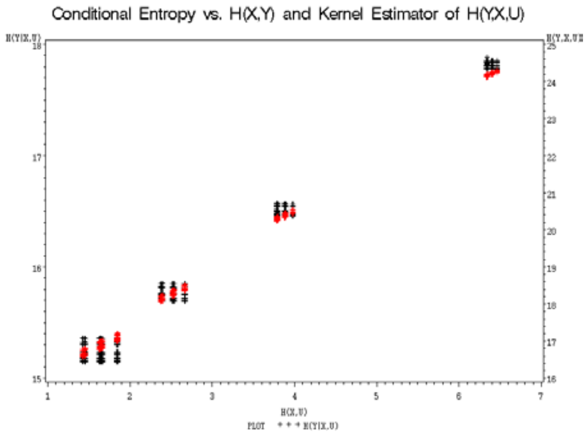

Figure 2.

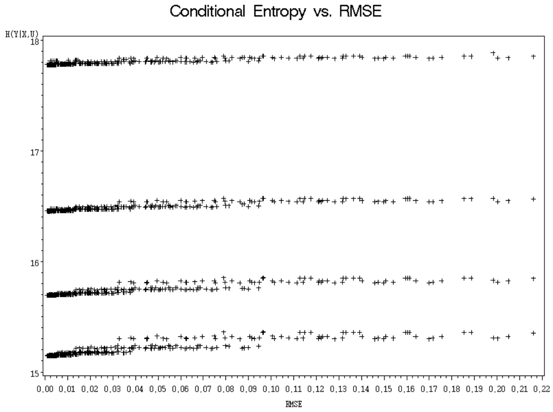

Figure 3.

I. Model M(2)

Simulation generates N subsmples of shape (y_{tk}, (x_t; u_t) , k = 1,…,M) , on which are calculated estimators of conditional entropies, by applying in the following way:

Step 1: generate X~Exp(λ_1)

Step 2: generate U~ Exp(λ_2)

Step 3: for each generated pair from Steps 1 and 2 generate M occurrences of Y as described in Section II.B.

Step 4: Calculate estimators of H(Y,X,U) and H(X,U)

Step5: Apply Proposition 4 and calculate kernel estimators for conditional entropies

Table III shows parameters of this simulation.

TABLE III.
SIMULATION PARAMETERS FOR M(2)

	M	N	λ_1	λ_2	σ
Range	10-1000	10-1000	[0,2] by 0.2	[0,2] by 0.2	1-3
	δ	β			
Range	1-5	1-5			

Figure4 shows relations between estimator of Conditional Entropy and Kernel estimator of H(Y,X,U) Figure 5 represents entropies as expressed in Proposition 4. Conditionality can be observed in both graphs.

Figure 6 shows that: the larger is Conditional Entropy the lower is RMSE.

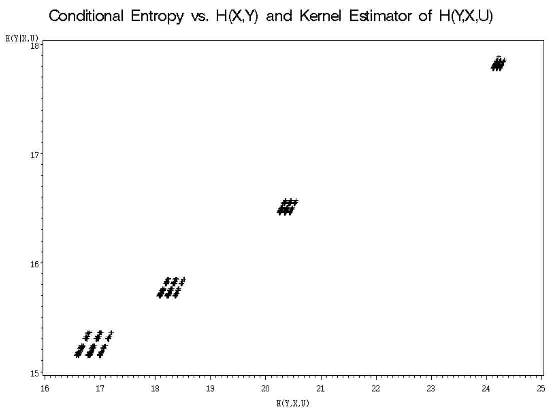

Figure 4.

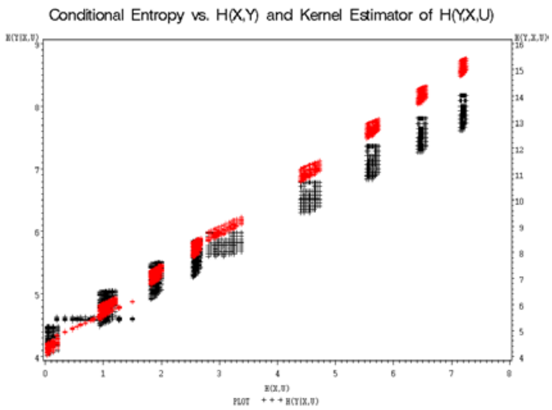

Figure 5.

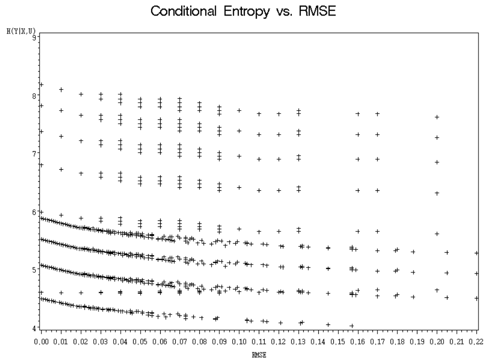

Figure 6.

Figure 7.

CONCLUSIONS

We define and study two conditional entropies *H (response* **|** *covariates)* for regression models and materialize them for logistic semiparametric regression models with binary response (LSpRModelsBR).

Based on its properties and on our simulation results, we conclude that a conditional entropy quantifies the intensity of the connection "response ~covariates" within a regression model. Therefore, we can identify a new goodness-of-fit index for regression models, as well as a new quantitative criterion for statistical modelling: *"The larger conditional entropy H (response* **|** *covariates), the better fitted the regression model response ~covariates".*

BIBLIOGRAPHY:

[1] Beirlant, J., Dudewicz , E.J., Gyor., L., Van der Meulen, E.C. (1997), Nonparametric Entropy Estimation. An overview, *International Journal of Mathematical and Statistical Sciences*, **6**(1)

[2] Cachin, C. (1997), Entropy measures and unconditional security in cryptography, Ph.D. dissertation, Dept. Comput. Inf. Sci., Swiss Federal Inst. Technol., Zürich, Switzerland

[3] Jizba P., Arimitsu, T.(2004), The world according to Rényi: Thermodynamics of multifractal systems, *Ann. Phys.*, 312, 17-59

[4] Golshani, L., Pasha, E., Yari, G. (2009), Some properties of Rényi entropy and Rényi entropy rate, *Inf. Sci.*, 179, 2426.2433

[5] Renner, R., Wolf, S., (Roy, B. Ed.), (2005), Simple and tight bounds for information reconciliation and privacy amplification, *Proc. 11th Int.Conf. Theory Appl. Cryptol. Inf. Security - ASIACRYPT*, 199 - 216.

[6] Xu, D., Erdogmuns, D. (2010), Renyi's entropy, divergence and their nonparametric estimators, published by Principe, J.C. (Ed), *Information Theoretic Learning*

[7] Teixeira, A., Matos, A., Antunes, L. (2012), Conditional Rényi Entropies. *IEEE Transactions on Information Theory*, 58(7), 4273 – 4277

[8] Rajagopalan, B., Lall, U. (1995), A kernel estimator for discrete distributions, *Nonparametric Statistics*, 4, 409 - 426

[9] Wang, M.C., Van Ryzin, J. (1981), A class of smooth estimators for discrete distributions, *Biometrika*, 68 (1), 301 - 309

[10] Shannon C.E. (1948), A mathematical theory of communication, *Bell Syst. Techn. Journal*, 27, 379-423, 623-656

[11] Renyi, A. (1961), On measures of information and entropy. Proc. of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, 547 – 561

[12] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode, *The Annals of Mathematical Statistics* **33** (3)