

A goodness-of-fit test of Student distributions based on Rényi entropy

Justine Lequesne

*Laboratoire de Mathématiques N. Oresme, UMR CNRS 6139, Campus 2, Université de Caen
Basse Normandie, BP 5186, 14032 Caen, France, justine.lequesne@unicaen.fr*

Abstract. The non-standard Student distributions maximize Rényi entropy among all distributions having a fixed variance. Based on this maximum entropy property, we construct a goodness-of-fit test for testing composite null hypotheses of the non-standard Student distributions. The proposed test thus generalizes the goodness-of-fit tests based on Shannon entropy introduced by O. Vasicek in 1976 for testing normality.

Keywords: goodness-of-fit tests, non-standard Student distributions, Rényi entropy.

INTRODUCTION

Introduced by Vasicek [1] for testing normality, goodness-of-fit tests based on Shannon entropy have been widely discussed in the literature. They are constructed for simple or composite null hypotheses for testing many different distributions in the exponential family. Their good asymptotic properties, simplicity of construction and efficiency in terms of power make them a good alternative to classical goodness-of-fit tests such as Kolmogorov-Smirnov or Cramér-von Mises. Mathematically, they are premised on maximum entropy properties of distributions in the exponential family; see Lequesne [2].

Extending the framework, we will construct a goodness-of-fit test for the Student distribution based on q -Rényi entropy. The Student distribution maximizes Rényi entropy among all distributions having a fixed variance, just as normal distributions maximize Shannon entropy. The test statistics also require a non-parametric estimator of Rényi entropy. A commonly used estimator is the nearest neighbors estimator; see Vasicek [1] and Leonenko *et. al.* [3] for construction and properties.

In the first section, the non-standard Student distribution is presented as a maximum q -Rényi entropy distribution. Then, the statistics for the goodness-of-fit test is constructed in the second section, with asymptotic properties given in the third. Computation of the critical values using Monte Carlo simulation is provided in the last section.

THE MAXIMUM RÉNYI ENTROPY DISTRIBUTIONS

Introduced by Rényi [4], the Rényi entropy of order $q > 0$ of any distribution P on \mathbb{R} – with density p with respect to the Lebesgue measure – is

$$\mathbb{H}_q(P) = \frac{1}{1-q} \log \left(\int p^q(x) dx \right).$$

The limiting value of \mathbb{H}_q as $q \rightarrow 1$ is Shannon entropy; in the following, we will restrict ourselves to the case $q \neq 1$.

The set $\mathcal{P}(\mathbf{m}, \mathbf{M})$ is a set of distributions satisfying moment constraints

$$C_j(\mathbf{m}, \mathbf{M}) = \mathbb{E}_P(M_j) = \int_{\mathbb{R}} M_j(x) p(x) dx = m_j, \quad (1)$$

where M_j are integrable functions with respect to P and $\mathbf{m} = (m_1, \dots, m_J) \in \mathbb{R}^J$ for some fixed $J \in \mathbb{N}$, and X is any real variable with distribution P . Distributions maximizing Shannon entropy in $\mathcal{P}(\mathbf{m}, \mathbf{M})$ are known to belong to the exponential family, with densities

$$p^*(x; \boldsymbol{\theta}) = Z(\boldsymbol{\theta}) \exp \left[\sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) M_j(x) \right], \quad (2)$$

where $\boldsymbol{\theta} \in \Theta \subset \overline{\mathbb{R}}^d$, with $d \leq J$ and $Z(\boldsymbol{\theta})$ is a normalizing constant; see Csiszár [5]. The exponential family contains many classical distributions, including Gaussian, exponential and uniform. The following similar result for Rényi entropy is proven in Grechuk *et al.* [6].

Theorem 1. *Let $\mathcal{P}(\mathbf{m}, \mathbf{M})$ be the set of distributions P satisfying moment constraints $C_j(\mathbf{m}, \mathbf{M})$ of the form (1). A unique parametric distribution*

$$P_q^*(\boldsymbol{\theta}) = \arg \max_{P \in \mathcal{P}(\mathbf{m}, \mathbf{M})} \mathbb{H}_q(P) \quad (3)$$

maximizes the q -Rényi entropy subject to linear constraints $C_j(\mathbf{m}, \mathbf{M})$ for $1 \leq j \leq J$. Its density is of the form

$$p_q^*(x; \boldsymbol{\theta}) = \left[-\frac{1}{q} \sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) M_j(x) \right]_+^{\frac{1}{q-1}}$$

$$= \begin{cases} \left(-\frac{1}{q} \sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) M_j(x) \right)^{\frac{1}{q-1}} & \text{if } \sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) M_j(x) \leq 0 \text{ or } q < 1, \\ 0 & \text{if } \sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) M_j(x) \geq 0 \text{ and } q > 1, \end{cases}$$

where $\lambda_j(\boldsymbol{\theta})$ are the Lagrange multipliers determined by the constraints.

The entropy of the maximum q -Rényi entropy distribution is

$$\mathbb{H}_q(P_q^*(\boldsymbol{\theta})) = \frac{1}{1-q} \log \left(-\frac{1}{q} \sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) m_j \right).$$

The family of the maximum q -Rényi entropy distributions is the q -exponential family of distributions with densities of the form

$$p_q(x; \boldsymbol{\theta}) = Z(\boldsymbol{\theta}) \exp_q \left[\sum_{j=1}^J \lambda_j(\boldsymbol{\theta}) M_j(x) \right],$$

where $\exp_q(x) = [1 + (1 - q)x]_+^{\frac{1}{1-q}}$ is called the q -exponential function. The q -exponential family contains many classical distributions such as the q -gaussian or the non-standard Student distributions.

The non-standard Student distribution, with location parameter μ , scale parameter σ^2 and ν degrees of freedom, that we will denote by $\mathcal{T}(\nu, \mu, \sigma^2)$, has a density on \mathbb{R} given by

$$f_\nu(x; \mu, \sigma^2) = \frac{1}{\sqrt{\nu\pi\sigma}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left[1 + \frac{1}{\nu} \left(\frac{x - \mu}{\sigma} \right)^2 \right]^{-\frac{\nu+1}{2}},$$

where $\Gamma(x)$ is the gamma function. Note that by a location-scale transformation, if a variable X has a $\mathcal{T}(\nu, \mu, \sigma^2)$ distribution, then $Y = (X - \mu)/\sigma$ has a classical Student distribution with ν degrees of freedom.

For $1/3 < q < 1$, the distribution maximizing the q -Rényi entropy among all distributions having the same mean μ and the same variance Σ^2 is the non-standard Student distribution $\mathcal{T}(\nu, \mu, \sigma^2)$ with

$$\nu = \frac{1+q}{1-q}, \quad (4)$$

and

$$\sigma^2 = \frac{\nu-2}{\nu} \Sigma^2, \quad (5)$$

where $\nu \in \mathbb{R}_+^*$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_+^*$. Since the q -Rényi entropy of this distribution depends only on q and the parameters σ^2 and ν , we will simply denote it by $h_q(\nu, \sigma^2)$. Precisely,

$$h_q(\nu, \sigma^2) = \frac{1}{1-q} \log \left[\frac{B\left(\frac{q(1+\nu)-1}{2}, \frac{1}{2}\right)}{B^q(\nu/2, 1/2)} \right] + \frac{1}{2} \log(\pi\nu\sigma^2) - \log\Gamma(1/2),$$

where $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ is the beta function.

For $q > 1$, the q -Rényi entropy maximizing distribution is a truncated version of the non-standard Student distribution, defined on a compact support. We will focus here on a goodness-of-fit test for the non-standard Student distributions defined on \mathbb{R} .

CONSTRUCTION OF THE STATISTICS

Let (X_1, \dots, X_n) be an independent and identically distributed sample of an unknown distribution P . We will construct a test of the composite hypothesis

$$H_0 : "P \in \mathcal{P}_0(\Theta)" \quad \text{against} \quad H_1 : "P \notin \mathcal{P}_0(\Theta)" \quad (6)$$

where $\mathcal{P}_0(\Theta) = \{\mathcal{T}(\nu, \mu, \sigma^2); \nu \in \mathbb{R}_+^*, \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*\}$ is the family of all non-standard Student distributions. It will be based on their maximum entropy property.

Testing a composite null hypothesis requires first estimating parameters. For the $\mathcal{F}(\nu, \mu, \sigma^2)$ distribution, two cases may be encountered, leading to two different test statistics.

If the null hypothesis is partially specified, ν is known and σ^2 is unknown, then σ^2 can be estimated using (5) by

$$\widehat{\sigma}_n^2 = \frac{\nu - 2}{\nu} \widehat{\Sigma}_n^2,$$

where $\widehat{\Sigma}_n^2$ is an estimator of the variance.

If the null hypothesis H_0 is totally unspecified, estimating both parameters ν and σ^2 is required. As a first approach, we propose to maximize the likelihood function of the Student distribution to obtain the maximum likelihood estimators (MLE). This method leads to implicit equations of which no explicit solutions can be determined; see Johnson and Kotz [7]. Alternatively, Liu and Rubin [8] propose to compute the MLE of the Student parameters via the EM algorithm.

Following the method developed by Vasicek [1] for testing normality, we propose the test statistic

$$\widehat{\mathbb{T}}_{n,q}^1 = h_q(\nu, \widehat{\sigma}_n^2) - \widehat{\mathbb{H}}_q(P)_n, \quad (7)$$

for a partially specified null hypothesis, or

$$\widehat{\mathbb{T}}_{n,q}^2 = h_q(\widehat{\nu}_n, \widehat{\sigma}_n^2) - \widehat{\mathbb{H}}_q(P)_n, \quad (8)$$

for a totally unspecified null hypothesis, where:

- $(\widehat{\nu}_n, \widehat{\sigma}_n^2)$ are MLE estimators of (ν, σ^2) ;
- $\widehat{\mathbb{H}}_q(P)_n$ is a non parametric estimator of the q -Rényi entropy of the unknown distribution P ;
- q is the order of the Rényi entropy, determined through (4) by $q = \frac{\nu-1}{\nu+1}$, where ν is replaced by $\widehat{\nu}_n$ when it is unknown.

In what follows, we simply denote by $\widehat{\mathbb{T}}_{n,q}^{1,2}$ both the two test statistics when they satisfy the same properties or converge similarly.

For estimating the q -Rényi entropy of any distribution P , Leonenko *et al.* [3] proposed an estimator based on the nearest-neighbor estimator of the unknown density p , and plug-in estimation of the q -moment of P defined by

$$I_q = \int_{\mathbb{R}} p^q(x) dx. \quad (9)$$

For a sample (X_1, \dots, X_n) , let $\rho_i^{(1)}$ be the nearest-neighbor Euclidean distance from X_i to X_j in the sample, with $j \neq i$, and similarly, let $\rho_i^{(k)}$ be the k -th nearest-neighbor distance from X_i to X_j . These distances form the order statistics $\rho_i^{(1)} \leq \rho_i^{(2)} \leq \dots \leq \rho_i^{(n-1)}$, and I_q is estimated by

$$\widehat{I}_{n,k,q} = \frac{1}{n} \sum_{i=1}^n (\zeta_{n,i,k})^{1-q},$$

where $\zeta_{n,i,k} = 2(n-1)G_k\rho_i^{(k)}$ with

$$G_k = \left[\frac{\Gamma(k)}{\Gamma(k+1-q)} \right]^{1/(1-q)}.$$

Finally, an estimator of the q -Rényi entropy is given by

$$\widehat{H}_{n,k,q} = \frac{\log(\widehat{I}_{n,k,q})}{1-q}. \quad (10)$$

ASYMPTOTIC PROPERTIES OF THE STATISTICS

Consistency properties of $\widehat{H}_{n,k,q}$ are proven in Leonenko *et al.* [3]. In particular, we state here the consistency of $\widehat{H}_{n,k,q}$ with $q < 1$ for distributions with unbounded supports, such as the non-standard Student distributions.

Proposition 1. *Let P be a distribution with density p with unbounded support. Let I_q be the q -moment of P defined in (9). Set*

$$r(p) = \sup\{r > 0 : \int_{\mathbb{R}} |x|^r p(x) dx < \infty\}.$$

If $I_q < \infty$ and $r(p) > (1-q)/q$, then

$$\mathbb{E}[\widehat{H}_{n,k,q}] - \mathbb{H}_q(P) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Moreover, if $r(p) > 2(1-q)/(2q-1)$, then $\mathbb{E}[\widehat{H}_{n,k,q} - \mathbb{H}_q(P)]^2$ converges to 0 when n tends to ∞ .

The consistency of $\widehat{H}_{n,k,q}$ yields consistency results for the test statistics $\widehat{\mathbb{T}}_{n,q}^{1,2}$, based on the convergence of the MLE estimators of parameters ν and σ^2 . To our knowledge, no reference exists in the literature that explicitly proves the latter. As an alternative, we have checked it through Monte Carlo simulation. For illustration, Figure 1 plots the computed MLE of the distribution $\mathcal{S}(5, 3, 0.8)$ for various sizes n of samples. Any other case may be dealt with similarly.

Theorem 2. *Let (X_1, \dots, X_n) be an n -sample drawn from an unknown distribution P . Suppose that $\mathbb{H}_q(P)$ is estimated by $\widehat{H}_{n,k,q}$ given in (10).*

For a goodness-of-fit test of (6), for which the non-standard Student distribution and P have the same unspecified variance Σ^2 , the following results hold true:

1. *For a partially specified null hypothesis with test statistics $\widehat{\mathbb{T}}_{n,q}^1$ given in (7), where $\widehat{\Sigma}_n^2$ is a consistent estimator of the variance, if $\nu > 4$, then under H_0 ,*

$$\widehat{\mathbb{T}}_{n,q}^1 \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty. \quad (11)$$

2. *For a totally unspecified null hypothesis with test statistics $\widehat{\mathbb{T}}_{n,q}^2$ given in (8), if $\widehat{\nu}_n > 4$, then under H_0 ,*

$$\widehat{\mathbb{T}}_{n,q}^2 \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty. \quad (12)$$

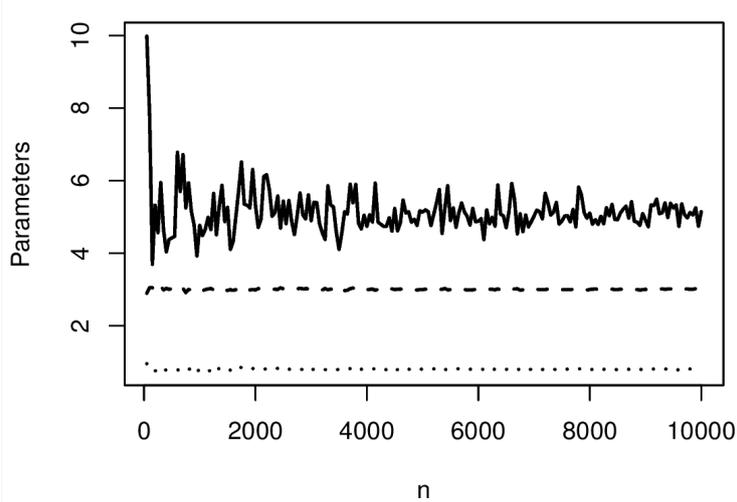


FIGURE 1. MLE of the non-standard Student distribution for various n . The continuous line corresponds to \widehat{v}_n , the dashed line to $\widehat{\mu}_n$ and the dotted line to $\widehat{\sigma}^2_n$.

3. Under the alternative hypothesis H_1 ,

$$\widehat{\mathbb{T}}_{n,q}^{1,2} \xrightarrow{P} C \quad \text{as } n \rightarrow \infty,$$

where C is a positive constant.

Proof. 1. The consistency of $\widehat{\Sigma}_n^2$ gives the consistency of $\widehat{\sigma}^2_n$ by (5). By the continuity of $h_q(v, \sigma^2)$ with respect to σ^2 , we get

$$h_q(v, \widehat{\sigma}^2_n) \xrightarrow{P} h_q(v, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

2. The continuity of $h_q(v, \sigma^2)$ with respect to v, σ^2 and q together with the convergence of the MLE estimators \widehat{v}_n and $\widehat{\sigma}^2_n$ yields the convergence

$$h_q(\widehat{v}_n, \widehat{\sigma}^2_n) \xrightarrow{P} h_q(v, \sigma^2) \quad \text{as } n \rightarrow \infty.$$

For proving Points 1 and 2, it remains to check that $\widehat{H}_{n,k,q}(v)$ is consistent. Conditions of Theorem 1 are satisfied: for $v > 4$ (or $\widehat{v}_n > 4$), the non-standard Student distribution has an unbounded support, and $r(\mathcal{T}(v, \mu, \sigma^2)) = v$, which satisfies

$$v > \frac{1-q}{q} \quad \text{and} \quad v > 2 \frac{1-q}{2q-1}.$$

Therefore, under H_0

$$\widehat{H}_{n,k,q} \xrightarrow{P} h_q(v, \sigma^2) \quad \text{as } n \rightarrow \infty,$$

and (11) and (12) hold true.

TABLE 1. Critical values of the tests.

	$n = 10$		$n = 20$		$n = 50$		$n = 100$	
$C^*(\alpha)^1$	$k = 1$	0.78	$k = 1$	0.54	$k = 1$	0.36	$k = 1$	0.25
	$k = 3$	0.63	$k = 4$	0.49	$k = 5$	0.33	$k = 5$	0.22
	$k = 5$	0.67	$k = 8$	0.56	$k = 10$	0.42	$k = 20$	0.37
	$k = 7$	0.59	$k = 16$	0.43	$k = 25$	0.49	$k = 50$	0.48
$C^*(\alpha)^2$	$k = 1$	0.88	$k = 1$	0.87	$k = 1$	0.58	$k = 1$	0.38
	$k = 3$	0.77	$k = 4$	0.98	$k = 5$	0.82	$k = 5$	0.46
	$k = 5$	0.82	$k = 8$	1.13	$k = 10$	1.05	$k = 20$	0.69
	$k = 7$	0.83	$k = 16$	1.03	$k = 25$	1.19	$k = 50$	0.89

3. Under the alternative hypothesis, for any distribution P such that $r(P)$ is sufficiently large, $\widehat{H}_{n,k,q}$ converges to $\mathbb{H}_q(P)$ as n tends to infinity. Thanks to the equality of variances, the maximum entropy constraint of the non-standard Student distribution is satisfied, which gives $h_q(v, \sigma^2) > \mathbb{H}_q(P)$. Therefore $\widehat{\mathbb{T}}_{n,q}^{1,2} \xrightarrow{P} C$ as $n \rightarrow \infty$, where C is a positive constant, and the result is proven. \square

COMPUTATION OF CRITICAL VALUES

The issue of a statical test is the rejection or non rejection of the null hypothesis by using a critical region. Thanks to the method of estimation of the parameters, P and the non-standard Student distribution have the same variance. By the maximum entropy property and Theorem 2, both test statistics $\widehat{\mathbb{T}}_{n,q}^{1,2}$ are positive and converge to zero under H_0 . The critical region is of the form $\mathcal{R}_\alpha = \{\widehat{\mathbb{T}}_{n,q}^{1,2} \leq C^*(\alpha)\}$ where $C^*(\alpha)$ is a strictly positive constant – called the critical value, depending on the significance level $\alpha \in (0, 1)$ of the test through

$$P_{H_0}(\widehat{\mathbb{T}}_n^{1,2} \leq C^*(\alpha)) = \alpha,$$

where P_{H_0} is the probability under the distribution of the null hypothesis. If the asymptotic distribution of the test statistics under the null hypothesis were known, $C^*(\alpha)$ would be the α -quantile of this distribution. Since the asymptotic distribution of $\widehat{H}_{n,k,q}$ and $\widehat{\mathbb{T}}_{n,q}^{1,2}$ are unknown, we have computed empirical quantiles through Monte Carlo simulation.

For various sample sizes n , $N = 3000$ observations of the distribution $\mathcal{T}(5, 2, 0.5^2)$ are simulated. The test statistics $\widehat{\mathbb{T}}_{n,q}^{1,2}$ are computed for all observations and several k . Then $C^*(\alpha)^{1,1}$ is the $(\alpha \times N)$ -th order statistic. Results are given in Table 1 for $\alpha = 0.05$. One can see that the critical values converge to zero slowly as n tends to infinity, and that the same holds for the test statistics. Finally, we suggest to choose for the estimator $\widehat{H}_{n,k,q}$ the k yielding the lower critical value, and thus the least conservative test.

FURTHER ISSUES

Goodness-of-fit tests for the non-standard Student distributions based on maximum q -Rényi properties are constructed above. Determining the asymptotic distribution for the test statistics, studying the power of the test, and also constructing of a faster

convergent estimator of $\mathbb{H}_q(P)$, are issues that remain to be investigated. Constructing an ad hoc estimator, instead of the MLE, for the maximum entropy method can be also considered. More generally, similarly to goodness-of-fit tests based on Shannon entropy for the exponential family, tests based on the q -Rényi entropy can be constructed for testing all distributions in the q -exponential family. Extension to other distributions and entropy functionals is also under current investigation. For more on these issues, see Lequesne [2].

REFERENCES

1. Vasicek O. A test for normality based on sample entropy. Journal of the Royal Statistical Society. Series B (Methodological), pages 54–59, 1976.
2. Lequesne J. Tests statistiques basés sur la théorie de l’information. applications en démographie et en biologie. PhD Thesis, 2015.
3. Leonenko N., Pronzato L., Savani V., et al. A class of Rényi information estimators for multidimensional densities. The Annals of Statistics, 36(5): pages 2153–2182, 2012.
4. Rényi A. On measures of entropy and information. In Fourth Berkeley Symposium on Mathematical Statistics and Probability, pages 547–561, 1961.
5. Csiszár I. I-divergence geometry of probability distributions and minimization problems. The Annals of Probability, pages 146–158, 1975.
6. Grechuk B., Molyboha A., and Zabaranin M. Maximum entropy principle with general deviation measures. Mathematics of Operations Research, 34(2): pages 445–467, 2009.
7. Johnson N. and Kotz S. Distributions in Statistics: Continuous Univariate Distributions-2. John Wiley and Sons, 1970.
8. Liu C. and Rubin D.B. ML estimation of the t distribution using EM and its extensions, ECM and ECME. Statistica Sinica, 5(1): pages 19–39, 1995.