Statistical inference for Rényi entropy

David Källberg

Department of Mathematics
and Mathematical Statistics
Umeå University

Coauthor
Oleg Seleznjev
Department of Mathematics
and Mathematical Statistics
Umeå University

Outline

- Introduction
- Measures of uncertainty
- U-statistics
- Estimation of entropy
- Numerical experiment
- Conclusion

Introduction

- A system described by probability distribution $\mathcal{P}$
- Only partial information about $\mathcal{P}$ is available, e.g. the mean
- A measure of uncertainty (entropy) in $\mathcal{P}$
- What $\mathcal{P}$ should we use if any?

Why entropy?

The entropy maximization principle: choose $\mathcal{P}$, satisfying given constraints, with maximum uncertainty.

Objectivity: we don't use more information than we have.

Measures of uncertainty. The Shannon entropy.

- Discrete $\mathcal{P} = \{p(k), k \in D\}$

$$h_1(\mathcal{P}) := -\sum_k p(k) \log p(k)$$

- Continuous $\mathcal{P}$ with density $p(x), x \in \mathbb{R}^d$

$$h_1(\mathcal{P}) := -\int_{\mathbb{R}^d} \log(p(x))p(x)dx$$

Measures of uncertainty. The Rényi entropy.

A class of entropies. Of order $s \neq 1$ given by

- Discrete $\mathcal{P} = \{p(k), k \in D\}$

$$h_s(\mathcal{P}) := \frac{1}{1-s} \log \big(\sum_k p(k)^s \big)$$

- Continuos $\mathcal{P}$ with density $p(x), x \in \mathbb{R}^d$

$$h_s(\mathcal{P}) := \frac{1}{1-s} \log \big(\int_{\mathbb{R}^d} p(x)^s dx \big)$$

Motivation

The Rényi entropy satisfies axioms on how a measure of uncertainty should behave, Rényi (1970).

For both discrete and continuous $\mathcal{P}$, the Rényi entropy is a generalization of the Shannon entropy, because

$$\lim_{q \to 1} h_q(\mathcal{P}) = h_1(\mathcal{P})$$

Problem

Non-parametric estimation of integer order Rényi entropy, for discrete and continuous $\mathcal{P}$, from sample $\{X_1, \ldots X_n\}$ of $\mathcal{P}$-iid observations.

Overview of Rényi entropy estimation, continuos $\mathcal{P}$

- Consistency of nearest neighbor estimators for any s, Leonenko *et al.* (2008)
- Consistency and asymptotic normality for quadratic case s=2, Leonenko and Seleznjev (2010)

U-statistics: basic setup

For a $\mathcal{P}$-iid sample $\{X_1, \ldots, X_n\}$ and a symmetric kernel function $h(x_1, \ldots, x_m)$

$$\mathbb{E}h(X_1, \ldots, X_m) = \theta(\mathcal{P})$$

The U-statistic estimator of θ is defined as:

$$U_n = U_n(h) := \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} h(X_{i_1}, \ldots, X_{i_m})$$

U-statistics: properties

- Symmetric, unbiased
- Optimality properties for large class of $\mathcal{P}$
- Asymptotically normally distributed

Estimation, continuous case

Method relies on estimating functional

$$q_s := \int_{\mathbb{R}^d} p^s(x)dx = \mathbb{E}(p^{s-1}(X))$$

Estimation, some notation

- $d(x, y)$ the Euclidean distance in $\mathbb{R}^d$
- $B_\epsilon(x) := \{y : d(x, y) \leq \epsilon\}$ ball of radius ϵ with center x.
- b_ϵ volume of $B_\epsilon(x)$
- $p_\epsilon(x) := P(X \in B_\epsilon(x))$ the ϵ-ball probability at x

Estimation, useful limit

When $p(x)$ bounded and continuous, we rewrite

$$q_s = \lim_{\epsilon \to 0} \mathbb{E}(p_\epsilon^{s-1}(X))/b_\epsilon^{s-1}$$

So, unbiased estimate of $q_{s,\epsilon} := \mathbb{E}(p_\epsilon^{s-1}(X))$ leads to asymptotically unbiased estimate of q_s as $\epsilon \to 0$.

Estimation of $q_{s,\epsilon}$

For $s = 2, 3, 4, \ldots$, let

$$I_{ij}(\epsilon) := I(d(X_i, X_j) \leq \epsilon)$$

$$\tilde{I}_i(\epsilon) := \prod_{\substack{1 \leq j \leq s \\ j \neq i}} I_{ij}(\epsilon)$$

Define U-statistic $Q_{s,n}$ for $q_{s,\epsilon}$ by kernel

$$h_s(x_1, \ldots, x_s) := \frac{1}{s} \sum_{i=1}^{s} \tilde{I}_i(\epsilon)$$

Estimation of Rényi entropy

Denote by $\tilde{Q}_{s,n} := Q_{s,n}/b_\epsilon^{s-1}$ an estimator of q_s and by $H_{s,n} := \frac{1}{1-s}\log\left(\max\left(\tilde{Q}_{s,n}, 1/n\right)\right)$ corresponding estimator of h_s

Consistency

Assume $\epsilon = \epsilon(n) \to 0$ as $n \to \infty$

Let $v_{s,n}^2 := Var(\tilde{Q}_{s,n})$

$v_{s,n}^2 \to 0$ as $n\epsilon^d \to a \in (0, \infty]$, so we get

Theorem

Let $n\epsilon^d \to a$, $0 < a \leq \infty$ and $p(x)$ be bounded and continuos. Then $H_{s,n}$ is a consistent estimator of h_s

Smoothness conditions

Denote by $H^\alpha(K)$, $0 < \alpha \leq 2$, $K > 0$, a linear space of continuos functions in $\mathbb{R}^d$ satisfying α-Hölder condition if $0 < \alpha \leq 1$ or if $1 < \alpha \leq 2$ with continuos partial derivates satisfying $(\alpha - 1)$-Hölder condition with constant K.

Asymptotic normality

When $n\epsilon^d \to \infty$, we have $v_{s,n}^2 \sim s(q_{2s-1} - q_s^2)/n$

Let $K_{s,n} = \max(s(\tilde{Q}_{2s-1,n} - \tilde{Q}_{s,n}^2), 1/n)$

$L(n) > 0, n \geq 1$ is a slowly varying function as $n \to \infty$

Theorem

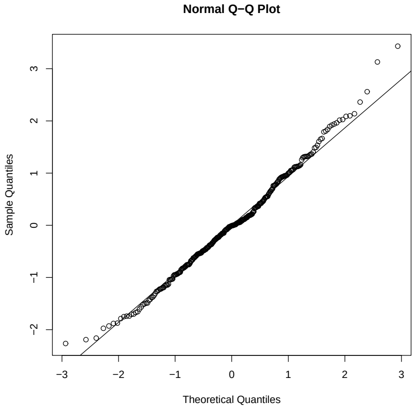
Let $p^{s-1}(x) \in H^\alpha(K)$ for $\alpha > d/2$. If $\epsilon \sim L(n)n^{-1/d}$ and $n\epsilon^d \to \infty$, then

$$\sqrt{n}\frac{\tilde{Q}_{s,n}(1-s)}{\sqrt{K_{s,n}}}(H_{s,n} - h_s) \xrightarrow{\mathcal{D}} N(0,1)$$

Numerical experiment

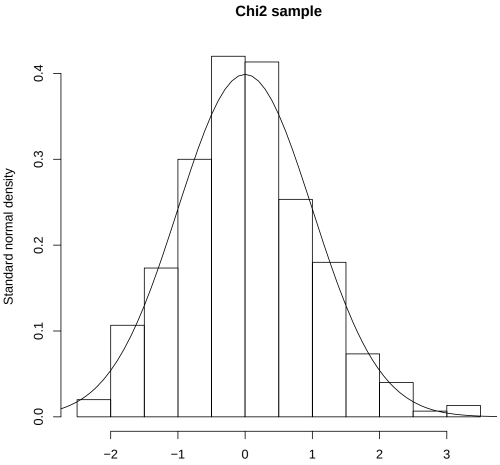
- χ^2 distribution, 4 degrees of freedom. $h_3 = -\frac{1}{2}\log(q_3)$, where $q_3 = 1/54$
- 300 simulations, each of size $\mathrm{n} = 500$.
- Quantile plot and histogram supports standard normality

Figures



Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−3
−2
−1
0
1
2
3

Figures



Conclusion

Asymptotically normal estimates possible for integer order Rényi entropy.

References

- Leonenko, N. ,Pronzato, L. ,Savani, V. (1982). A class of Rényi information estimators for multidimensional densities, *Annals of Statistics* **36** 2153-2182
- Leonenko, N. and Seleznjev, O. (2009). Statistical inference for ϵ-entropy and quadratic Rényi entropy. Univ. Umeå, Research Rep., Dep. Math. and Math. Stat., 1-21, *J. Multivariate Analysis* (submitted)
- Renyi, A. *Probability theory,* North-Holland Publishing Company **1970**