Statistical inference for Rényi entropy of integer order

David Källberg

August 23, 2010

Outline

- Introduction
- Measures of uncertainty
- Estimation of entropy
- Numerical experiment
- More results
- Conclusion

Coauthor:

Oleg Seleznjev
Department of Mathematics and Mathematical Statistics

Umeå university

Introduction

- A system described by probability distribution $\mathcal{P}$.
- Only partial information about $\mathcal{P}$ is available, e.g. the covariance matrix.
- 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 $p(x), x \in \mathbb{R}^d$

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

A class of entropies. Of order $s \geq 0$ given by

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

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

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

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

Motivation

The Rényi entropy satisfies axioms on how a measure of uncertainty should behave, Rényi(1961,1970).

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

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

Problem

Non-parametric estimation of integer order Rényientropy, for discrete and continuous multivariate $\mathcal{P}$, from sample $\{X_1, \ldots X_n\}$ of $\mathcal{P}$-i.i.d. observations.

Estimators of entropy are widely used.

- Distribution identification problems (Student-r distributions).
- Average case analysis for random databases.
- Clustering.

Overview of Rényi entropy estimation

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

Notation

- $I(C)$ indicator function of event C.
- $\mathcal{S}_{s,n}$ set of all s-subsets of $\{1, \ldots, n\}$.
- $\sum_{(s,n)}$ is summation over $\mathcal{S}_{s,n}$.
- $d(x, y)$ 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

Method relies on estimating functional

$$q_s := \mathbb{E}(p^{s-1}(X)) = \begin{cases} \sum_k p(k)^s & (\mathit{Discrete}) \\ \int_{\mathbb{R}^d} p(x)^s dx & (\mathit{Continuous}) \end{cases}$$

Estimation, discrete case

An estimator of q_s,

$$Q_{s,n} := \binom{s}{n}^{-1} \sum_{(s,n)} \psi(S),$$

where

$$\psi(S) := \frac{1}{s} \sum_{i \in S} I(X_i = X_j, \forall j \in S).$$

- $H_{s,n} := \frac{1}{1-s} \log(\max(\tilde{Q}_{s,n}, 1/n))$ estimator of h_s.
- $Q_{s,n}$ is a U-statistic, so properties follows from conventional theory.

Estimation, continuous case

A U-statistic estimator of $q_{\epsilon,s} := \mathbb{E}(p_\epsilon(X)^{s-1})$,

$$Q_{s,n} := \binom{s}{n}^{-1} \sum_{(s,n)} \psi(S),$$

where

$$\psi(S) := \frac{1}{s} \sum_{i \in S} I(d(X_i, X_j) \leq \epsilon, \forall j \in S).$$

- $\tilde{Q}_{s,n} := Q_{s,n}/b_\epsilon(d)^{s-1}$ asymptotically unbiased estimator of q_s if $\epsilon = \epsilon(n) \to 0$ as $n \to \infty$.
- $H_{s,n} := \frac{1}{1-s} \log(\max(\tilde{Q}_{s,n}, 1/n))$ estimator of h_s.

Smoothness conditions

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

Asymptotics, continuous case

- $\epsilon = \epsilon(n) \to 0$ as $n \to \infty$.
- $L(n) > 0, n \geq 1$, a slowly varying function as $n \to \infty$.
- $K_{s,n} := \max(s^2(\tilde{Q}_{2s-1,n} - \tilde{Q}^2_{s,n}), 1/n)$ consistent estimator of the asymptotic variance.

Theorem

Suppose that $p(x)$ is bounded and continuous. Let $n\epsilon^d \to a$ for some $0 < a \leq \infty$.

(i) *Then $H_{s,n}$ is a consistent estimator of h_s.*

(ii) *Let $p(x)^{s-1} \in H^{(\alpha)}(K)$ for some $d/2 < \alpha \leq 2$. If $\epsilon \sim L(n)n^{-1/d}$ and $a = \infty$, then*

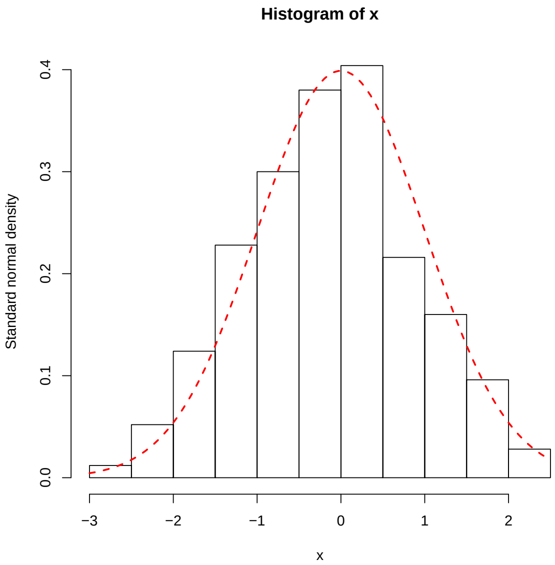
$$\sqrt{n}\frac{\tilde{Q}_{s,n}(1-s)}{\sqrt{K_{s,n}}}(H_{s,n} - h_s) \xrightarrow{\text{D}} N(0,1) \quad \text{as} \quad n \to \infty.$$

Numerical experiment

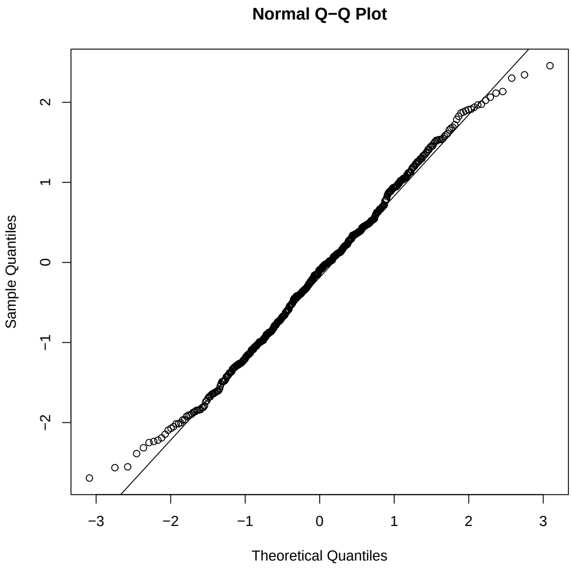
- χ^2 distribution, 4 degrees of freedom. $h_3 = -\frac{1}{2}\log(q_3)$, where $q_3 = 1/54$.
- 500 simulations, each of size $\mathrm{n} = 1000$, $\epsilon = 1/4$.
- Quantile plot and histogram supports standard normality.

Remark. Choice of ϵ in a practical situation remains an open problem.

Figures



Histogram of x
Standard normal density
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
x


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

More results

- Estimation of (quadratic) Rényi entropy from m-dependent sample.

- Two different distributions $\mathcal{P}_X$ and $\mathcal{P}_Y$. Inference for

  ...functionals of type

$$\int_{R^d} p_X(x)^{s_1} p_Y(x)^{s_2} dx, \quad s_1, s_2 \in \mathbb{N}_+.$$

  ...statistical distances (Bregman)

  - Discrete:

$$D_2 := \sum_k (p_X(k) - p_Y(k))^2$$

  - Continuous:

$$D_2 := \int_{R^d} (p_X(x) - p_Y(x))^2 dx$$

Conclusion

Asymptotically normal estimates possible for Rényi entropy of integer order.

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. (2010). Statistical inference for ϵ-entropy and quadratic Rényientropy, *J. Multivariate Analysis* **101**, Issue 9, 1981-1994.
- Rényi, A.(1961). On measures of information and entropy *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 547-561.
- Renyi, A. *Probability theory,* North-Holland Publishing Company **1970**