Asymptotic behaviour of the weighted Shannon differential entropy in a Bayesian problem

Mark Kelbert, Pavel Mozgunov

2nd International Electronic Conference on Entropy and Its Applications

November 2015

Introduction

Let $\mathrm{U} \sim \mathbb{U}[0,1]$.
Given a realization of this RV p, consider a sequence of conditionally independent identically distributed $(\xi_i, i = 1, 2, \ldots)$, where $\xi_i = 1$ with probability p and $\xi_i = 0$ with probability $1 - p$. Let x_i, each 0 or 1, be an outcome in trial i.
Denote $S_n = \xi_i + \ldots + \xi_n$ and $x = \sum_{i=1}^{n} x_i$.

$$\mathbb{P}(\xi_i = 1, \xi_j = 1) = \int_0^1 p^2 \mathrm{d}p = 1/3 \ \text{ if } i \neq j, \ \text{ but}$$

$$\mathbb{P}(\xi_i = 1)\mathbb{P}(\xi_j = 1) = (\int_0^1 p \mathrm{d}p)^2 = 1/4.$$

The probability that after n trials the exact sequence $(x_i,\ i = 1, \ldots, n)$ will appear equals

$$\mathbb{P}(\xi_1 = x_1, \ldots, \xi_n = x_n) = \int_0^1 p^x (1-p)^{n-x} \mathrm{d}p = \frac{1}{(n+1)\binom{n}{x}}. \qquad (1)$$

Introduction

This implies that the posterior probability density function (PDF) of the number of x successes after n trials is uniform

$$\mathbb{P}(S_n = x) = \frac{1}{(n+1)}, x = 0, \ldots, n.$$

The posterior PDF given the information that after n trials one observes x successes takes the form

$$f^{(n)}(p|\xi_1 = x_1, ..., \xi_n = x_n) = f^{(n)}(p|S_n = x) = (n+1)\binom{n}{x} p^x (1-p)^{n-x}, \tag{2}$$

Note that conditional distribution given in (2) is a **Beta-distribution**.

"It is known that Beta-distribution is *asymptotically normal* with its mean and variance as x and $(n-x)$ tend to infinity, but this fact is *lacking a handy reference*"

Introduction

Consider RV $Z^{(n)}$ on $[0;1]$ with PDF (2). Note that $Z^{(n)}$ has the followings expectation:

$$\mathbb{E}_x[Z^{(n)}] = \frac{x+1}{n+2}, \tag{3}$$

and the following variance:

$$\mathbb{V}_x[Z^{(n)}] = \frac{(x+1)(n-x+1)}{(n+3)(n+2)^2}. \tag{4}$$

Shannon's differential entropy

The goal of our previous work [13] was to study the asymptotic behaviour of the differential entropy (DE) of the following RVs:

1. $Z_\alpha^{(n)}$ with PDF $f_\alpha^{(n)}$ given in (2) when $x = x(n) = \lfloor \alpha n \rfloor$ where $0 < \alpha < 1$ and $\lfloor a \rfloor$ is integer part of a.
2. $Z_\beta^{(n)}$ with PDF $f_\beta^{(n)}$ given in (2) when $x = x(n) = \lfloor n^\beta \rfloor$ where $0 < \beta < 1$
3. $Z_{c_1}^{(n)}$ with PDF $f_{c_1}^{(n)}$ given in (2) when $x = c_1$ and $Z_{n-c_2}^{(n)}$ with PDF $f_{n-c_2}^{(n)}$ given in (2) when $n - x(n) = c_2$ where c_1 and c_2 are some constants.

It is shown that in the **first** and **second** cases limiting distribution is *Gaussian* and the differential entropy of standardized RV converges to differential entropy of *standard Gaussian RV*.

In the **third** case the limiting distribution in *not Gaussian*, but still the asymptotic of differential entropy can be found explicitly.

Recall

Differential entropy (DE) $h(f)$ of a RV Z with the PDF f:

$$h(f) = h_{diff}(f) = -\int_{\mathbb{R}} f(z)\log(f(z))\mathrm{d}z \tag{5}$$

with the convention $0\log 0 = 0$.

A linear transformation $X = b_1 Z + b_2$,

$$h(g) = h(f) + \log b_1 \tag{6}$$

where g is a PDF of RV X.

Let $\bar{Z}$ be the standard Gaussian RV with PDF φ then the differential entropy of $\bar{Z}$ equals:

$$h(\varphi) = \frac{1}{2}\log(2\pi e).$$

Recall the definition of the Kullback–Leibler divergence of g from f

$$\mathbb{D}(f||g) = \int_{\mathbb{R}} f(x)\log\frac{f(x)}{g(x)}\mathrm{d}x. \tag{7}$$

Shannon's differential entropy. Case I

Theorem

Let $\tilde{Z}_\alpha^{(n)} = n^{\frac{1}{2}}(\alpha(1-\alpha))^{-\frac{1}{2}}(Z_\alpha^{(n)} - \alpha)$ *be a RV with PDF* $\tilde{f}_\alpha^{(n)}$*. Let* $\bar{Z} \sim \mathcal{N}(0,1)$ *be the standard Gaussian RV, then*

(a) $\tilde{Z}_\alpha^{(n)}$ *weakly converges to* $\bar{Z}$*:*

$$\tilde{Z}_\alpha^{(n)} \Rightarrow \bar{Z} \text{ as } n \to \infty.$$

(b) *The differential entropy of* $\tilde{Z}_\alpha^{(n)}$ *converges to differential entropy of* $\bar{Z}$*:*

$$\lim_{n\to\infty} h(\tilde{f}_\alpha^{(n)}) = \frac{1}{2}\log(2\pi e).$$

(c) *The Kullback-Leibler divergence of* φ *from* $\tilde{f}_\alpha^{(n)}$ *tends to* 0 *as* $n \to \infty$*:*

$$\lim_{n\to\infty} \mathbb{D}(\tilde{f}_\alpha^{(n)} || \varphi) = 0.$$

Shannon's differential entropy. Case I

We obtained the following asymptotic of the differential entropy:

$$\lim_{n\to\infty}\left[h(f_\alpha^{(n)}) - \frac{1}{2}\log\frac{2\pi e[x(n-x)]}{n^3}\right] = 0. \quad (8)$$

Particularly,

$$\lim_{n\to\infty}\left[h(f_\alpha^{(n)}) - \frac{1}{2}\log\frac{2\pi e[\alpha(1-\alpha)]}{n}\right] = 0. \quad (9)$$

Due to (6), the differential entropy of RV $\tilde{Z}_\alpha^{(n)}$ has the form:

$$\lim_{n\to\infty}\left[h(\tilde{f}_\alpha^{(n)})\right] = \frac{1}{2}\log(2\pi e). \quad (10)$$

Shannon's differential entropy. Case II

Theorem

Let $\tilde{Z}_\beta^{(n)} = n^{1-\beta/2}(Z_\beta^{(n)} - n^{\beta-1})$ *be a RV with PDF* $\tilde{f}_\beta^{(n)}$ *and* $\bar{Z} \sim \mathcal{N}(0,1)$ *then*

(a) $\tilde{Z}_\beta^{(n)}$ *weakly converges to* $\bar{Z}$*:*

$$\tilde{Z}_\beta^{(n)} \Rightarrow \bar{Z} \text{ as } n \to \infty.$$

(b) *The differential entropy of* $\tilde{Z}_\beta^{(n)}$ *converges to differential entropy of* $\bar{Z}$*:*

$$\lim_{n\to\infty} h(\tilde{f}_\beta^{(n)}) = \frac{1}{2}\log(2\pi e).$$

(c) *The Kullback-Leibler divergence of* φ *from* $\tilde{f}_\beta^{(n)}$ *tends to* 0 *as* $n \to \infty$*:*

$$\lim_{n\to\infty} \mathbb{D}(\tilde{f}_\beta^{(n)} || \varphi) = 0.$$

Shannon's differential entropy. Case III

Theorem

Let $\tilde{Z}_{c_1}^{(n)} = nZ_{c_1}^{(n)}$ *be a RV with PDF* $\tilde{f}_{c_1}^{(n)}$ *and* $\tilde{Z}_{n-c_2}^{(n)} = nZ_{n-c_2}^{(n)}$ *be a RV with PDF* $\tilde{f}_{n-c_2}^{(n)}$*. Denote* $H_k = 1 + \frac{1}{2} + \ldots + \frac{1}{k}$ *the partial sum of harmonic series and* γ *the Euler-Mascheroni constant, then*

(a) $$\lim_{n\to\infty} h(\tilde{f}_{c_1}^{(n)}) = c_1 + \sum_{i=0}^{c_1-1} \log(c_1 - i) - c_1(H_{c_1} - \gamma) + 1.$$

(b) $$\lim_{n\to\infty} h(\tilde{f}_{n-c_2}^{(n)}) = c_2 + \sum_{i=0}^{c_2-1} \log(c_2 - i) - c_2(H_{c_2} - \gamma) + 1.$$

Motivation of the weighted differential entropy

Consider the following statistical experiment with **twofold goal**:

1. **on the initial stage** an experimenter is mainly concerns whether the coin is approximately **fair** with a high precision.
2. **As the size of a sample grows**, he proceeds to estimate **the true value** of the parameter anyway.

We want to *quantify the differential entropy of this experiment* taking into account its two sided objective.

Quantitative measure of information gain of this experiment is provided by the concept of

the weighted differential entropy.

Introducing the weight function

Let $\phi^{(n)} \equiv \phi^{(n)}(\alpha, \gamma, p)$ be a weight function that underlines the importance of some particular value γ.

Choosing the weight function we adopt the following normalization rule:

$$\int_{\mathbb{R}} \phi^{(n)} f_{\alpha}^{(n)} \mathrm{d}p = 1. \tag{11}$$

Weighted differential entropies

The goal of the this work is to study the asymptotic behaviour of weighted Shannon's (12) and Renyi's differential entropies of RV $Z^{(n)}$ with PDF $f^{(n)}$ given in (2) and particular RV $Z_\alpha^{(n)}$ with PDF $f_\alpha^{(n)}$ given in (2) with $x = \lfloor \alpha n \rfloor$ where $0 < \alpha < 1$:

$$h^\phi(f_\alpha^{(n)}) = -\int_{\mathbb{R}} \phi^{(n)} f_\alpha^{(n)} \log f_\alpha^{(n)} \mathrm{d}p, \tag{12}$$

$$H_\nu^\phi(f_\alpha^{(n)}) = \frac{1}{1-\nu} \log \int_{\mathbb{R}} \phi^{(n)} \left(f_\alpha^{(n)} \right)^\nu \mathrm{d}p \tag{13}$$

where $\nu \geq 0$ and $\nu \neq 1$.

The weight function $\phi^{(n)}$

The following special cases are considered:

1. $\phi^{(n)} \equiv 1$
2. $\phi^{(n)}$ depends both on n and p

In this paper we consider the weight function of the following form:

$$\phi^{(n)}(p) = \Lambda^{(n)}(\alpha, \gamma) p^{\gamma\sqrt{n}} (1-p)^{(1-\gamma)\sqrt{n}} \tag{14}$$

where $\Lambda^{(n)}(\alpha, \gamma, p)$ is found from the normalizing condition (11).

This is the model example with a **twofold goal**:

- **to emphasize a particular value** γ (for moderate n)
- **asymptotically unbiased** estimate

$$\lim_{n\to\infty} \int_0^1 p\phi^{(n)} f^{(n)} \mathrm{d}p = \alpha.$$

The weighted Shannon differential entropy

Theorem

For the weighted Shannon differential entropy of RV $Z_\alpha^{(n)}$ with PDF $f_\alpha^{(n)}$ and weight function $\phi^{(n)}$ given in (14) the following limit exists

$$\lim_{n\to\infty}\left(h^\phi(f_\alpha^{(n)})-\frac{1}{2}\log\left(\frac{2\pi e\alpha(1-\alpha)}{n}\right)\right)=\frac{(\alpha-\gamma)^2}{2\alpha(1-\alpha)}. \qquad (15)$$

If the $\alpha=\gamma$ then

$$\lim_{n\to\infty}\left(h^\phi(f_\alpha^{(n)})-h(f_\alpha^{(n)})\right)=0 \qquad (16)$$

where $h(f_\alpha^{(n)})$ is the standard ($\phi\equiv 1$) Shannon's differential entropy.

The weighted Shannon differential entropy

The normalizing constant in the weight function (14) is found from the condition (11). We obtain that

$$\Lambda^{(n)}(\gamma) = \frac{\Gamma(x+1)\Gamma(n-x+1)\Gamma(n+2+\sqrt{n})}{\Gamma(x+\gamma\sqrt{n}+1)\Gamma(n-x+1+\sqrt{n}-\gamma\sqrt{n})\Gamma(n+2)} = \frac{\mathbb{B}(x+1, n-x+1)}{\mathbb{B}(x+\gamma\sqrt{n}+1, n-x+\sqrt{n}-\gamma\sqrt{n}+1)} \tag{17}$$

where $\Gamma(x)$ is the Gamma function and $\mathbb{B}(x, y)$ is the Beta function. We denote by $\psi^{(0)}(x) = \psi(x)$ and by $\psi^{(1)}(x)$ the digamma function and its first derivative respectively.

Recall the Stirling formula:

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right) \text{ as } n \to \infty. \tag{18}$$

The weighted Renyi differential entropy

Theorem

Let $Z^{(n)}$ be a RV with PDF $f^{(n)}$ given in (2), $Z_\alpha^{(n)}$ be a RV with PDF $f_\alpha^{(n)}$ given in (2) with $x = \lfloor \alpha n \rfloor$, $0 < \alpha < 1$ and $H_\nu(f^{(n)})$ be the weighted Renyi differential entropy given in (13).

(a) *When $\phi^{(n)} \equiv 1$ and both (x) and $(n - x)$ tend to infinity as $n \to \infty$ the following limit holds*

$$\lim_{n\to\infty} \left(H_\nu(f^{(n)}) - \frac{1}{2}\log\frac{2\pi x(n-x)}{n^3} \right) = -\frac{\log(\nu)}{2(1-\nu)}, \tag{19}$$

(b) *When the weight function $\phi^{(n)}$ is given in (14) the following limit for the Renyi weighted entropy of $f_\alpha^{(n)}$ holds*

$$\lim_{n\to\infty} \left(H_\nu^\phi(f_\alpha^{(n)}) - \frac{1}{2}\log\frac{2\pi\alpha(1-\alpha)}{n} \right) = -\frac{\log(\nu)}{2(1-\nu)} + \frac{(\alpha-\gamma)^2}{2\alpha(1-\alpha)\nu}, \tag{20}$$

The weighted Renyi differential entropy

$$H_\nu(f^{(n)}) = \frac{1}{2}\log\left(\frac{2\pi x(n-x)}{n^3}\right) - \frac{\log(\nu)}{2(1-\nu)} + O\left(\frac{1}{n}\right). \qquad (21)$$

Note that the leading terms in (21) looks like Renyi differential entropy of Gaussian RV with variance $\sigma^2 = \frac{x(n-x)}{n^3}$.

Taking the limit when $\nu \to 1$ and applying L'Hopital's rule we get that

$$H_{\nu\to1}(f^{(n)}) = \lim_{\nu\to1} H_\nu(f^{(n)}) = \frac{1}{2}\log\left(\frac{2e\pi x(n-x)}{n^3}\right) + O\left(\frac{1}{n}\right). \qquad (22)$$

For example, when $x = \lfloor \alpha n \rfloor$, $0 < \alpha < 1$ the Renyi entropy:

$$H_{\nu\to1}(f^{(n)}) = \frac{1}{2}\log\frac{2\pi e[\alpha(1-\alpha)]}{n} + O\left(\frac{1}{n}\right)$$

where the leading term is Shannon's differential entropy of Gaussian RV with corresponding variance.

The weighted Renyi differential entropy

Theorem

For any continuous random variable X with PDF f and for any non-negative weight function $\phi(x)$ which satisfies condition (11) and such that

$$\int_{\mathbb{R}} \phi(x) f(x)^{\nu} |\log(f(x))| \mathrm{d}x < \infty,$$

the weighted Renyi differential entropy $H_{\nu}^{\phi}(f)$ is a non-increasing function of ν and

$$\frac{\partial}{\partial \nu} H_{\nu}^{\phi}(f) = -\frac{1}{(1-\nu)^2} \int_{\mathbb{R}} z(x) \log \frac{z(x)}{\phi(x) f(x)} dx \tag{23}$$

where

$$z(x) = \frac{\phi(x)(f(x))^{\nu}}{\int_{\mathbb{R}} \phi(x)(f(x))^{\nu} \mathrm{d}x}$$

Further extension

Natural extension of this work is to derive **the weighted analogous of the Fisher Information** and the generalized version of well-known inequalities for the weighted variance

- Cramér-Rao inequality
- Bhattacharyya inequality

and for weighted Kullback distance

- Kullback inequality

Similar models of sensitive estimator appear in many fields of statistics. So, application of the weighted differential entropy approach can be adapted to a large variety of problems.

Thank you for attention!

M. Belis, S. Guiasu, *A quantitative and qualitative measure of information in cybernetic systems* (1968), IEEE Trans. Inf. Th.,14, 593-594

A. Clim, *Weighted entropy with application*, Analele Universitatii Bucurestica Matematica (2008), Anul LVII, 223-231.

T.M. Cover, J.M. Thomas, *Elements of Information Theory*, NY: Basic Books (2006)

G. Dial, I.Taneja, *On weighted entropy of type (α, β) and its generalizations.*, Appl. Math., 1981

R.L. Dobrushin, *Passing to the limit under the sign of the information and entropy*, Th. Prob. Appl., (1960), 29-37

I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Product* (2007), Elsevier

S. Guiasu, *Weighted entropy*, Report on Math. Physics, 2 (1971), 165–179.

J. Kapur., *Measures of Information and Their Applications.*, Chapter 17, New Delhi: Wiley Eastern Limited, 1994.

M. Kelbert, Yu. Suhov, *Continuity of mutual entropy in the large signal-to-noise ratio limit*, Stochastic Analysis (2010), Berlin: Springer, 281–299

M. Kelbert, Yu. Suhov, *Information Theory and Coding by Example*, Cambridge: Cambridge University Press, 2013

M. Kelbert, P. Mozgunov, *Shannon's differential entropy asymptotic analysis in a Bayesian problem* (2015), Mathematical Communications Vol 20, No 2 (2015)(in press)

M. Kelbert, Yu.Suhov, S.Y. Sekeh, *Weighted Fisher Information inequality* (2015), arXiv

M. Kelbert, P. Mozgunov, *Asymptotic behaviour of weighted differential entropies in a Bayesian problem* (2015), arXiv 1504.01612

Yu. Suhov, S. Yasaei Sekeh, M. Kelbert, *Entropy-power inequality for weighted entropy*, arXiv:1502.02188, 2015

Yu. Suhov, S. Yasaei Sekeh, *Simple inequalities for weighted entropies*, arXiv:1409.4102, 2015